From INSTITUTE OF ENVIRONMENTAL MEDICINE
Karolinska Institutet, Stockholm, Sweden

THE IMPACT OF TRANSCRIPTIONAL MUTAGENESIS ON CELLULAR HOMEOSTASIS

Monika Ezerskyte

Karolinska Institutet

Stockholm 2018

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB

ISBN 978-91-7831-241-2

The Impact of Transcriptional Mutagenesis on Cellular Homeostasis

THESIS FOR DOCTORAL DEGREE (Ph.D.)

To be publicly defended at Atrium, Nobels väg 12B, Solna

Friday the 7th of December 2018, at 09:30

By

Monika Ezerskyte

Principal Supervisor:
Docent Kristian Dreij
Karolinska Institutet
Institute of Environmental Medicine
Unit of Biochemical Toxicology

Co-supervisor(s):
Professor Ulla Stenius
Karolinska Institutet
Institute of Environmental Medicine
Unit of Biochemical Toxicology

Docent Magnus Olsson
Karolinska Institutet
Department of Environmental Medicine
Unit of Biochemical Toxicology

Opponent:
Professor Wim Vermeulen
Erasmus University Medical Center
Department of Molecular Genetics

Examination Board:
Professor Galina Selivanova
Karolinska Institutet
Department of Microbiology,
Tumor and Cell Biology

Docent Marianne Farnebo
Karolinska Institutet
Department of Bioscience and Nutrition

Professor Neus Visa
Stockholm University
Department of Molecular Biosciences,
The Wenner-Gren Institute

Per aspera ad astra

ABSTRACT

DNA is exposed to chemical and physical agents that represent a continuous threat to its integrity by formation of DNA lesions. Despite the existence and efficiency of DNA repair systems in cells, some lesions may not be removed, and interfere not only with the fidelity of DNA replication but also with DNA transcription. In fact, RNA polymerase can bypass non-bulky lesions, such as 8-oxoguanine (8-oxoG) and O^6-methylguanine (O^6-meG) and, due to base misincorporation, lead to the production of mutated RNA through a process referred to as transcriptional mutagenesis (TM). Although the concept of TM is well known, the biological consequences are still a recent discovery. While the few existing studies of TM have begun to shed light on the process, the role these seemingly transient errors might play in disease processes, especially in tumorigenesis, is currently unknown. The overall aim of this thesis was to explore to which extent lesion-induced TM influences protein function and how this may affect cellular homeostasis *in vivo*. Using plasmids containing single site-specific DNA lesions placed within probe genes with established links between specific mutations and subsequent phenotypes, we confirmed that 8-oxoG and O^6-meG induced TM in mammalian cells. Further, we explored the effects of TM on splicing fidelity *in vivo* (Paper I) and found that TM in regulatory sequences of splicing signals resulted in activation of alternative splicing sites, thus leading to the production of disease associated splice forms and/or disrupting physiological ratios between alternatively spliced isoforms. In addition, we examined effects of TM on p53 and its tumor suppressor function in human cells (Paper II). We found that expression of mutant R248W p53 due to TM was sufficient to reduce p53's transactivation capacity of several target genes, which are required for its tumor suppressive function. Moreover, we showed that TM of p53 reduced its tumor suppressor function by impairing both proper cell cycle control and induction of apoptosis resulting in stimulated proliferation and survival. A genome-wide gene expression analysis further revealed that TM of p53 at codon 248 deregulated both the transactivation and downregulation of numerous target genes, which are crucial for its tumor suppressor function (Paper III). These deregulated genes were involved in regulation of several cellular processes, such as cell-cycle arrest, apoptosis, and DNA damage response. To conclude, the work presented here shed light on biological effects of TM *in vivo* and provides evidence for possible mechanisms by which TM might contribute to human disease development. We showed for the first time that lesion-induced TM could activate alternative splicing sites *in vivo*, thus reducing splicing fidelity and resulting in aberrant splicing. In addition, the work presented in this thesis, together with the results from other studies, strongly suggest that TM could be a contributing mechanism in the multistep process of tumorigenesis by inactivating a tumor suppressor or activating an oncogene thus stimulating proliferation and survival of an already initiated pre-neoplastic cell.

LIST OF SCIENTIFIC PAPERS

I. Paredes JA, Ezerskyte M, Bottai M, Dreij K. Transcriptional mutagenesis reduces splicing fidelity in mammalian cells. *Nucleic Acids Research* 45(11): 6520–6529 (2017).

II. Ezerskyte M, Paredes JA, Malvezzi S, Burns JA, Margison GP, Olsson M, Scicchitano DA, and Dreij K. O^6-methylguanine–induced transcriptional mutagenesis reduces p53 tumor-suppressor function. *PNAS* 115(18): 4731–4736 (2018).

III. Ezerskyte M, Wang J, Pelechano V, Dreij K. Transcriptional mutagenesis dramatically alters genome-wide p53 transactivation landscape (*Manuscript*).

CONTENTS

1 INTRODUCTION....1
1.1 Eukaryotic transcription....1
1.1.1 Phases of transcription....2
1.1.2 mRNA splicing....4
1.2 Fidelity of transcription....7
1.3 DNA damage and transcription....9
1.3.1 Transcription-coupled repair....10
1.4 Transcriptional mutagenesis....12
1.5 Introduction to the study....14
1.5.1 How to study transcriptional mutagenesis....14
1.5.2 DNA damages and their repair....15
1.5.3 Reporter proteins....17
2 AIM OF THE STUDY....21
3 METHODS AND CONSIDERATIONS....23
3.1 Construction of damaged plasmids....23
3.1.1 Minigene reporter constructs....23
3.1.2 p53 TM plasmid....25
3.2 Cell lines....26
3.3 Transfection....27
3.4 Flow cytometry....28
3.5 Quantification of TM....29
3.6 RNA sequencing....30
3.7 Cell cycle analysis....31
3.8 Gene and protein expression....32
3.8.1 Real time qPCR and reverse transcription PCR....32
3.8.2 Western blot....32
3.9 Mitochondrial membrane potential....33
3.10 Caspase-3/-7-like activity assay....33
3.11 AGT activity assay....34
3.12 Cell viability assay....34
4 RESULTS....35
4.1 Paper I: Transcriptional Mutagenesis reduces splicing fidelity in mammalian cells....35
4.2 Paper II: O^6-methylguanine-induced Trancriptional Mutagenesis reduces p53 tumor-suppressor function....37
4.3 Paper III: Trancriptional Mutagenesis dramatically alters genome-wide p53 transactivation landscape....39
5 GENERAL DISCUSSION....41
6 CONCLUSIONS AND FUTURE OUTLOOK....47
7 ACKNOWLEDGEMENTS....49
8 REFERENCES....51

LIST OF ABBREVIATIONS

8-oxoG	8-oxoguanine
AGT	O^6-alkylguanine-DNA alkyltransferase
ATCC	American Type Culture Collection
BrdU	5-bromo-2'-deoxyuridine
CCCP	carbonyl cyanide 3-chlorophenylhydrazone
DEG	differentially expressed gene
DMEM	Dulbecco's Modified Eagle Medium
dNTP	deoxynucleoside triphosphate
GFP	green fluorescent protein
GO	gene ontology
HCG	high-confidence gene
HGPS	Hutchinson-Gilford progeria syndrome
IPA	Ingenuity Pathway Analysis
LMNA	lamin A
MEF	mouse embryonic fibroblasts
MEM	Minimum Essential Medium
MMP ($\Delta\Psi m$)	mitochondrial membrane potential
NTP	nucleoside triphosphate
O^6-bzG	O^6-benzylguanine
O^6-meG	O^6-methylguanine
OGG1	8-oxoguanine DNA glycosylase
PI	propidium iodide
PLP1	proteolipid protein 1
PMD	Pelizaeus–Merzbacher disease
RFP	red fluorescent protein
RNA Pol	RNA polymerase
TM	transcriptional mutagenesis
TMRE	tetramethylrhodamine, ethyl ester

1 INTRODUCTION

Homeostasis can be defined as the property of cells, tissues, and organisms to maintain and regulate stable and constant internal conditions required for a proper function when dealing with external changes. The word *homeostasis*, literally meaning "staying the same", derives from the Greek with words *homoios* meaning "similar" and *stasis* meaning "standing still" or "stable". At the cellular level, an intricate network of well-orchestrated regulatory mechanisms maintains the balance. Components of the network monitor changes of the environment, both internal and external, integrating a variety of signals through modulation of gene expression to adjust multiple cellular processes enabling a proper response to the stimulus. Gradual loss of proper homeostasis is a reason of many diseases and aging (Hartl, 2016).

The core of homeostasis maintenance is termed as the concept of the central dogma of molecular biology (introduced by Crick in 1958), which describes that the genetic information stored in DNA is transcribed into complementary RNA transcripts, which are subsequently translated into proteins (Figure 1). After discovery of reverse transcriptases and RNA viruses this concept has been extended to also include that information can be transferred from RNA to DNA and from RNA to RNA.

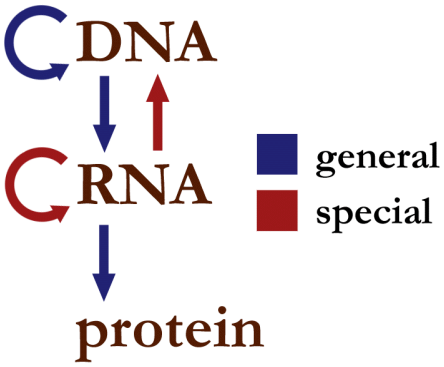

Figure 1. Flow of genetic information in biological systems –the central dogma of molecular biology. By Narayanese at English Wikipedia - Own work, Public Domain.

Transcription of genetic information into complementary RNA is a fundamental cellular process that characterizes identity and phenotype of a cell. This process can be modified by environmental factors, and thus impact on physiological processes within the cell. So, as for DNA replication, the regulation and fidelity of transcription is essential for proper maintenance of cellular homeostasis and the survival of all living organisms.

1.1 EUKARYOTIC TRANSCRIPTION

Each genome, a long sequence of DNA, contains all of the information needed to build and maintain the organism that carries it. The information stored in the genome is decoded into molecules of complementary RNA by RNA polymerases (RNA Pols) during a process called transcription that occurs in the nucleus of a cell (Lewin et al., 2011). The messenger RNAs (mRNAs), produced from protein-coding genes, are used as the templates for ribosomes to

instruct protein synthesis, whereas other non-coding RNAs, such as transfer RNA (tRNA), ribosomal (rRNA), microRNA, and small interfering RNA (siRNA), possess structural, regulatory, and catalytic functions. Transcription is the first event of gene expression and must therefore be regulated accurately and effectively to ensure proper homeostasis of all living organisms (von Hippel, 1998).

RNA Pols are conserved in all organisms and exhibit remarkable structural similarities (Ebright, 2000). These enzymes are large multisubunit protein complexes that catalyze synthesis of different RNAs. In contrast to single RNA Pol-containing prokaryotes, three different types of RNA Pols, named RNA Pol I, II, and III, exist in eukaryotic cells, but each is responsible for transcription of different type of genes. RNA Pol I transcribes the genes encoding most of the rRNA, whereas RNA Pol III transcribes DNA to synthesize ribosomal 5S RNA, tRNA that play a key role in the translation process, and some small nuclear regulatory RNA (snRNA). RNA Pol II synthesizes precursors of mRNA, most of snRNA, microRNA, and siRNA (Lewin et al., 2011). RNA Pol II is the most studied eukaryotic RNA Pol and is the focus of the thesis. Therefore, only the mammalian enzyme and its mRNA product transcribed from protein coding genes are further reviewed.

1.1.1 Phases of transcription

It is needed to mention that before eukaryotic transcription can begin and proceed, the chromatin has to be in an open configuration for general transcription factors to access the core promoter of the transcribed gene, which is accomplished by the recruitment of chromatin-remodeling complexes. Moreover, nucleosomes has to be removed from the promoter and displaced or relocated as RNA Pol II moves along the gene by mechanisms that are not fully resolved (Kulaeva et al., 2013). However, this section gives a brief overview of the current eukaryotic transcription model of RNA Pol II without focus on the impact of chromatin.

Transcription is a multistep process typically divided into distinct phases, initiation, promoter clearance, elongation, and termination.

Transcription initiation requires the binding of general transcription factors (GTFs), named TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH, to the core promoter in a defined order recruiting RNA Pol II at the transcription start site (TSS), resulting in the assembly of the closed preinitiation complex (PIC) (Orphanides et al., 1996). The RNA Pol II core promoters are structurally and functionally diverse regulatory sequences, but share several common sequence motifs, including an Initiator (Inr) region that encompasses the TSS, a TATA box which is usually located ~25 bp upstream of the TSS, and downstream promoter elements (DPEs) (Juven-Gershon and Kadonaga, 2010). TFIID, which consist of TATA binding protein (TBP) and a dozen of TAF (*TBP-Associated Factor*) subunits, recognizes and binds to the core promoter via TATA box recruiting components of the PIC (Burley and Roeder, 1996). Recognition of TATA-less core promoters for PIC assembly is mediated by the TAF subunits. In the generally accepted model of the PIC assembly, TFIIA and TFIIB

simultaneously bind next to further stabilize the interaction between TFIID and promoter DNA, which is followed by recruitment of RNA Pol II likely associated with TFIIF (Nogales et al., 2017; Roeder, 1996). Finally, formation of a transcriptionally competent PIC is completed by positioning of TFIIE and TFIIH, of which the latter is responsible for DNA melting. Then, TFIIH, using its ATP-dependent DNA helicase activity, causes ~10 bp region of duplex DNA to unwind upstream of the TSS allowing access to single-strand DNA template and resulting in the formation of an open initiation complex referred to as the "transcription bubble"(Kim et al., 2000).

Before RNA Pol II can enter a productive elongation phase, it must complete a phase known as a promoter clearance. The mechanism for the transition from transcriptional initiation to elongation phase is not fully elucidated, but several important sequential processes have been clarified. Once the open initiation complex is formed, RNA pol II catalyzes the phosphodiester bond formation between the first few nucleoside triphosphates (NTP) complementary to the template strand sequence yielding an initial RNA product (Lewin et al., 2011). When a ~10-12 nt transcript is synthesized, the 5′-end is separated from the template strand and enters the RNA exit channel stabilizing the transcription complex and the promoter clearance can occur (Luse, 2013; Sims et al., 2004). At the same time, phosphorylation of the RNA Pol II carboxyl-terminal repeat domain (CTD) is activated further inducing the promoter clearance (Svejstrup, 2004). Further phosphorylation of the CTD allows for recruitment of elongation stimulating factors, such as Spt4/5, Elf1, and TFIIS, to form a stable processive elongation complex (EC) that drives transcription elongation (Ehara et al., 2017).

At elongation phase, the EC reads the template (non-coding) DNA strand in 3′→5′ direction catalyzing the addition of new nucleotides via phosphodiester bond formation to the 3′-end of the growing transcript. Elongation rates of RNA Pol II are not constant and vary not only throughout the gene, but also between genes ranging ~1-6 kb per minute (Jonkers and Lis, 2015). A key rate-limiting step for transcription is pausing of EC 30-60 nt downstream of the TSS that is potentially subject to regulatory control. Moreover, the elongation process is tightly coupled to mRNA processing, including pre-mRNA capping, splicing (see *Section 1.1.2*), and even 3′-end polyadenylation (Proudfoot et al., 2002).

The last phase of the transcription process, termination, occurs when RNA Pol II dissociates from the DNA template downstream of the 3′-end of the transcript. RNA Pol II terminates at varying distances from the 3′-end of the mature mRNA and is directly coupled to 3′-end processing. The 3′-ends of mRNAs are generated by cleavage and subsequent synthesis of poly(A) tail to protect the transcript from degradation. The proper 3′-end processing reaction depends on multi-subunit protein complexes associated with the EC. For simplicity, *cleavage stimulatory factor* (CstF) complex recognizes consensus poly(A) motif and triggers the *cleavage and polyadenylation specific factor* (CPSF) complex to cleave the mRNA, whereas *poly(A) polymerase* (PAP) subsequently synthesizes poly(A) (Lewin et al., 2011; Richard and Manley, 2009).

The present model of transcription termination is a combination of allosteric/anti-terminator (Logan et al., 1987) and torpedo models (Connelly and Manley, 1988) and is described below (Richard and Manley, 2009). The model explains how the release of a full-length mature transcript triggers transcription termination by RNA Pol II. At the end of the transcribed gene, transcription past consensus poly(A) motif (AAUAAA) and the subsequent RNA cleavage ~10-30 nt downstream the poly(A) motif lead to the conformational changes of the EC. These changes may induce shedding off some of the elongation factors while recruiting termination factors. After cleavage, the resulting uncapped 5′-end of the RNA can be bound by specific exonuclease (Xrn2) that degrade it faster than the still bound RNA is being transcribed. Upon "catching up" with RNA Pol II, Xrn2 establishes interactions with RNA Pol II and triggers the dissociation of RNA pol II causing transcription to terminate (Lewin et al., 2011).

1.1.2 mRNA splicing

Pre-mRNA splicing is an essential step in the expression of nearly all eukaryotic genes. The vast majority of protein coding genes in all genomes from yeast to humans (and some prokaryotes) are discontinuous, containing a varying number of coding exons interrupted by noncoding sequences, called introns. Pre-mRNA splicing is a conserved process, in which introns are removed and flanking exons are joined together to form a mature mRNA, and occurs simultaneously with transcription (Brody and Shav-Tal, 2011; Proudfoot et al., 2002). Coupling splicing to transcription allows not only for temporal RNA processing before the completion of transcription, but also enables sequential recognition of RNA processing signals on nascent transcripts providing another means for gene expression regulation (Pandit et al., 2008).

Splicing is a tightly regulated process and requires extreme precision that highly depends on proper recognition of exons that can be buried within a sea of nucleotides. Various mutations either in consensus sequences of splice sites or auxiliary regulatory elements can disrupt both splicing and alternative splicing resulting in aberrantly spliced mRNA transcripts with deleterious consequences to the organism. Indeed, a broad range of human diseases is caused by mutations affecting splicing (Cooper et al., 2009; Ward and Cooper, 2010). Different types of RNA splicing, such as splicing executed by spliceosome, self-splicing introns or ribozymes, which are capable to catalyze their own excision from precursor RNA, tRNA splicing, and recursive splicing, occur in organisms. In this section, a metazoan splicing model of the major spliceosome is reviewed.

1.1.2.1 Splicing mechanism

The basic mechanism of splicing has been studied in detail and is well characterized (Figure 2). The process of splicing is dependent on special sequence motifs within introns that are recognized by different components of the large splicing apparatus, named spliceosome (described below). These sequences, called splicing signals, are the 5′ splice site, branch point and highly conserved polypyrimidine tract and 3′ splice site (5′ SS, 3′ SS, BPS and Py in

Figure 2) (Lewin et al., 2011; Sharp, 1987). The excision of introns from pre-mRNA occurs at exon-intron junction boundaries called splice sites. For the majority of mammalian introns, the 5′ splice site at the 5′ end of the intron is defined by a well-conserved GU dinucleotide encompassed within a less conserved sequence. The 3′ end region of the intron contains several splicing signals (Moore, 2000). The branch point, which always contains an adenosine at a proper position, but otherwise is less conserved, lies 18-40 nt upstream 3′ splice site followed by the highly conserved polypyrimidine tract. A terminal AG dinucleotide at the 3′ end of the intron defines 3′ splice site. The important role of the branch site is to form an RNA lariat structure that identifies the nearest 3′ splice site for targeted exon joining (Reed and Maniatis, 1988).

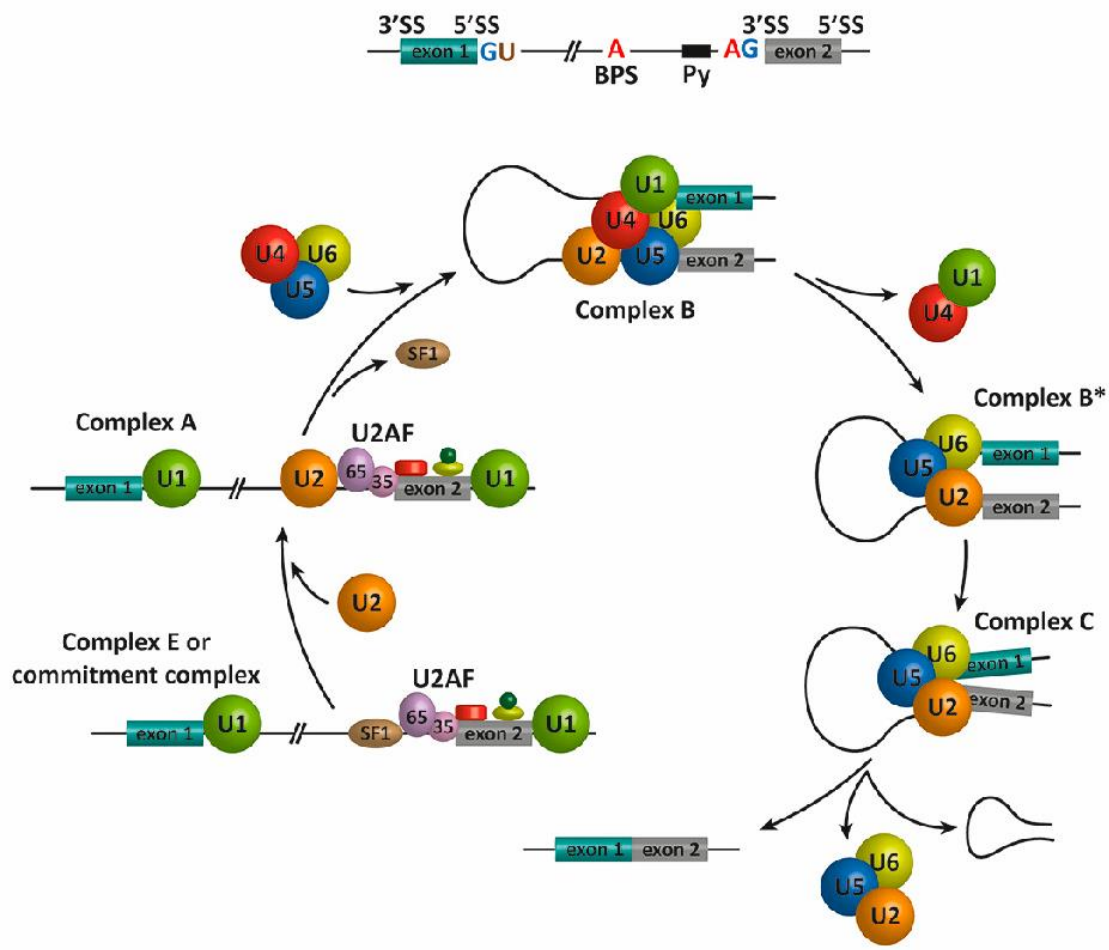

Figure 2. Schematic representation of spliceosome assembly and pre-messenger RNA splicing. Image reused from Suñé-Pou et al, 2017, with permission from Genes under Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

The spliceosome is a large multi-component splicing machinery consisting of both proteins and RNAs. The spliceosome that removes the vast majority of introns contains five small nuclear RNAs (snRNAs) associated with proteins in their natural state forming small ribonucleoproteins (snRNPs, pronounced as snurps), named U1, U2, U5, U4 and U6, and hundreds of other supplementary proteins (Lewin et al., 2011; Sharp, 1987). Assembly of the catalytically active spliceosome onto the pre-mRNA initiated by the recognition of splice

sites within the intron is a dynamic process requiring a series of structural rearrangements coordinated by multiple RNA–RNA, RNA–protein and protein–protein interactions.

Briefly, assembly of the spliceosome begins when U1 snRNP is recruited at 5′ splice site via base pairing between splice site and U1 snRNA and 3′ splice site is bound by a special set of proteins simultaneously (Figure 2). Thus the first relatively stable complex, named commitment complex or E complex, is formed. Next, in an ATP-dependent manner, U2 snRNP binds the branch site (involving base pairing between the sequence in U2 snRNA and the branch point) to form pre-spliceosome complex (A complex). The A complex is subsequently joined by a trimer containing U5 and U4/U6 snRNP converting the pre-spliceosome to a mature spliceosome (B complex). Finally, a series of complicated rearrangements result in the establishment of an activated spliceosome with a catalytic site for the splicing reaction to take place (Hastings and Krainer, 2001; Rino and Carmo-Fonseca, 2009; Schellenberg et al., 2008).

Once the activated spliceosome is formed, the splicing of pre-mRNA occurs in a two-step sequential reaction (Complex C, Figure 2) (Chiara and Reed, 1995). During the first reaction, the conserved adenosine within the branch site initiates a nucleophilic attack on the 5′ splice site. As a result, the upstream exon is released as a linear molecule and a branched or lariat splicing intermediate is formed through a new phosphodiester bond establishment. In the second reaction, the newly formed 3′ end of the exon attacks the 3′ splice site phosphodiester liberating the intron as a lariat and covalently joining the exons. (Lewin et al., 2011).

1.1.2.2 Alternative splicing

The architecture of the interrupted gene assists the process of alternative splicing to generate various combinations of mature mRNA including or excluding particular exons and even introns from an individual protein coding gene (Black, 2003). The basic patterns of alternative splicing include *e.g.* exon skipping, intron retention, alternative 5′and 3′splice site selection, and mutually exclusive exons, all of which coordinate the differential exclusion or inclusion of sequences of pre-mRNA into the mature mRNA and thus generating multiple mRNA variants (Black, 2003). In addition to basic patterns, many complex alternative splicing patterns also exist in the transcriptome (Vaquero-Garcia et al., 2016). Moreover, it is a major mechanism by which complex organisms regulate protein expression and dramatically increase the diversity of the transcriptome and proteome than what would be expected from the human genome containing roughly 20,000-25,000 genes (Black, 2003; Nilsen and Graveley, 2010; Stein, 2004). High-throughput RNA sequencing led to the estimation that ~95% of human multi-exon genes undergo alternative splicing (Pan et al., 2008; Wang et al., 2008). The resulting mRNA variants from alternative splicing can have divergent properties affecting gene expression in the cells (Matlin et al., 2005). Most alternative splicing events affect the coding sequences of mRNA resulting in the expression of different protein isoforms, which may have related, distinct or even opposing biological functions, *e.g.* different isoforms of Fas receptor produced by alternative splicing have opposing effects on apoptosis (Cascino et al., 1995). In addition, alternative splicing can also

change properties of mRNA itself, such as stability and localization, by introducing or removing important regulatory sequences (Lewin et al., 2011).

Short consensus splicing signals within the pre-mRNA that function to direct the splicing machinery can be separated by long sequences, therefore it is not surprising that additional sequence elements outside of the splice sites exist and can strongly affect pre-mRNA splicing. Indeed, both introns and exons contain additional *cis*-acting elements, such as intronic or exonic splicing enhancers (ISE or ESE) or silencers (ISS or ESS), that strongly affect assembly of spliceosome. These elements can recruit a large number of specific RNA-binding *trans*-acting splicing factors (Wang and Burge, 2008) that positively or negatively modulate the splice site selection affecting the outcome. Alternative splicing and splicing in general is a tightly regulated and extremely accurate process that can be specific to a cell type, developmental stage or even signaling pathway (Fox-Walsh and Hertel, 2009). Selection of splice site and differential use of most exons are orchestrated by a dynamic interplay between positive and negative *cis*-acting RNA elements and *trans*-acting regulators as well as the relative functional strength of splice site (House and Lynch, 2008; Matlin et al., 2005).

1.2 FIDELITY OF TRANSCRIPTION

Even though information transfer during replication, transcription and translation is conducted with remarkable precision, errors are inevitable and can have deleterious consequences for the organism especially when they affect protein-coding sequences. The majority of research has focused on the impact of replicative mutations to determine the importance of replication fidelity for organismal homeostasis. As a result, the contributions of replicative errors to various diseases, especially cancer, are well established (Bielas and Loeb, 2005; Hanahan and Weinberg, 2011; Loeb and Loeb, 2000). However, the impact of transcriptional errors has received much less attention than replication fidelity because transcription errors are transient and inheritable, remains hard to detect and quantify due to need of technically challenging and specialized methods. Consequently, transcription error rates (spontaneous, or in the absence of known DNA lesions) in different organisms as well as the mechanism of RNA Pol fidelity are still poorly investigated. The rate of transcription errors has so far been reported in RNA viruses (Acevedo et al., 2014), *E. coli* bacteria (Traverse and Ochman, 2016), *S. cerevisiae* yeast (Gout et al., 2017; Reid-Bayliss and Loeb, 2017; Shaw et al., 2002) and *C. elegans* nematodes (Gout et al., 2013) and ranges between 10^{-5}–10^{-6} per nucleotide, but is yet to be determined in higher organisms. In contrast, replicative fidelity has been well studied and characterized in a variety of species and environmental conditions. DNA Pol error rates have been roughly estimated to be 10^{-8}–10^{-10} per base per generation in various species and tissues (Bielas and Loeb, 2005; Lynch, 2010).

In addition, little is known about the RNA Pol II subunits that contribute to transcription fidelity. Only a few clues about mechanisms that ensure transcription fidelity have been revealed. RNA Pol fidelity is achieved through a stepwise process and involves two major

strategies: a correct substrate selection via induced fit mechanism and an intrinsic proofreading that involves the recognition and removal of a misincorporated nucleotide (Gamba and Zenkin, 2018; Sydow and Cramer, 2009). RNA Pol II selects correct NTP substrate firstly by discriminating NTPs from dNTPs by the recognition of 2′-OH group of the ribose. Secondly, NTPs that correctly base pair with the base on the DNA template strand and induce folding of the Trigger loop are incorporated into the transcript. The Trigger loop is a flexible domain of the RNA Pol active center and is required for the catalysis of phosphodiester bond formation (Yuzenkova et al., 2010).

Despite these mechanisms to ensure the selection of the correct substrate, misincorporations still occur. In the case of misincorporation, the intrinsic proofreading is initiated. The absence of proper base pairing with the template, the DNA:RNA heteroduplex can be destabilized which impairs the catalysis of the subsequent nucleoside monophosphate. Removal of the mismatched nucleotide involves backtracking of RNA Pol II and cleavage by an intrinsic 3′→5′ nuclease activity, resulting in a new RNA 3′-OH group at the active site, allowing RNA synthesis to resume. Several lines of evidence have identified that RNA Pol II subunit Rpb9 may affect transcriptional proofreading and thereby fidelity (Knippa and Peterson, 2013; Nesser et al., 2006). Additionally, various organisms possess factors that assist proofreading of transcription by stimulating hydrolysis of incorrect base, *e.g.* TFIIS for RNA Pol II (Gamba and Zenkin, 2018). Proofreading mechanisms are one major contributor to the overall transcription fidelity; however, backtracking induces pauses to elongating RNA Pol which poses other threats to the cells, such as conflicts with replication forks (Dutta et al., 2011).

As shown above, transcription error rates are several orders of magnitude higher compared to replication error rates, thus alterations in RNA sequence are also expected to produce transcripts that encode misfolded or malfunctioning proteins, although the production might be transient compared to fixed mutations in the genome. However, each mRNA molecule can be translated multiple times (Schwanhausser et al., 2011) resulting in further amplification of the error. As a result, a large pool of misfolded and malfunctioning proteins can burden the cells with potentially harmful consequences. Still very little is known about how transcriptional errors may affect cellular phenotypes and homeostasis, especially in higher eukaryotes, and about their possible contributions to disease development. Indeed, several studies in various organisms have revealed that transcription errors may affect protein function consequently leading to profoundly altered cellular phenotypes. Transcription errors can generate abnormal proteins in patients with non-familial Alzheimer's disease and Down syndrome (van Leeuwen et al., 1998), accelerate cellular aging in yeast (Vermulst et al., 2015), or even contribute to oncogenic pathways in mammalian cells (Brulliard et al., 2007). In recent years, next-generation sequencing approaches have allowed studies of genome-wide transcription fidelity (Carey, 2015; Gout et al., 2013; Gout et al., 2017; Reid-Bayliss and Loeb, 2017; Traverse and Ochman, 2016) that confirmed adverse effects of transcriptional errors on protein function and cellular homeostasis (Gout et al., 2017). These studies revealed a previously unappreciated role for

transcriptional fidelity in cellular homeostasis presenting a new mechanism by which cells can acquire disease phenotypes (Gordon et al., 2015; Vermulst et al., 2015). To date there are no genome-wide studies on transcription errors in mammalian cells, thus the extent of transcription errors as wells as the consequences of these events for cellular homeostasis in higher organisms are still unknown.

1.3 DNA DAMAGE AND TRANSCRIPTION

Cells transcribe thousands of genes simultaneously to direct synthesis of proteins required for maintenance of normal physiological processes. Even a small temporal distortion in gene expression can thus have detrimental consequences. As motivated above, RNA Pols require undamaged DNA templates, to produce correct and functional RNA transcripts. However, and in addition to inherent properties of fidelity, various exogenous and endogenous agents continuously damage DNA and pose a constant threat to the genome integrity. Thus, encounters of elongating RNA Pol II with DNA lesions in the template strands of expressed genes most likely occur frequently. This encounter has two possible outcomes: arrest of RNA Pol II elongation, which further can induce several distinct responses (Svejstrup, 2002), or bypass of RNA Pol II with a correct or incorrect nucleotide insertion event (Figure 3) (Doetsch, 2002).

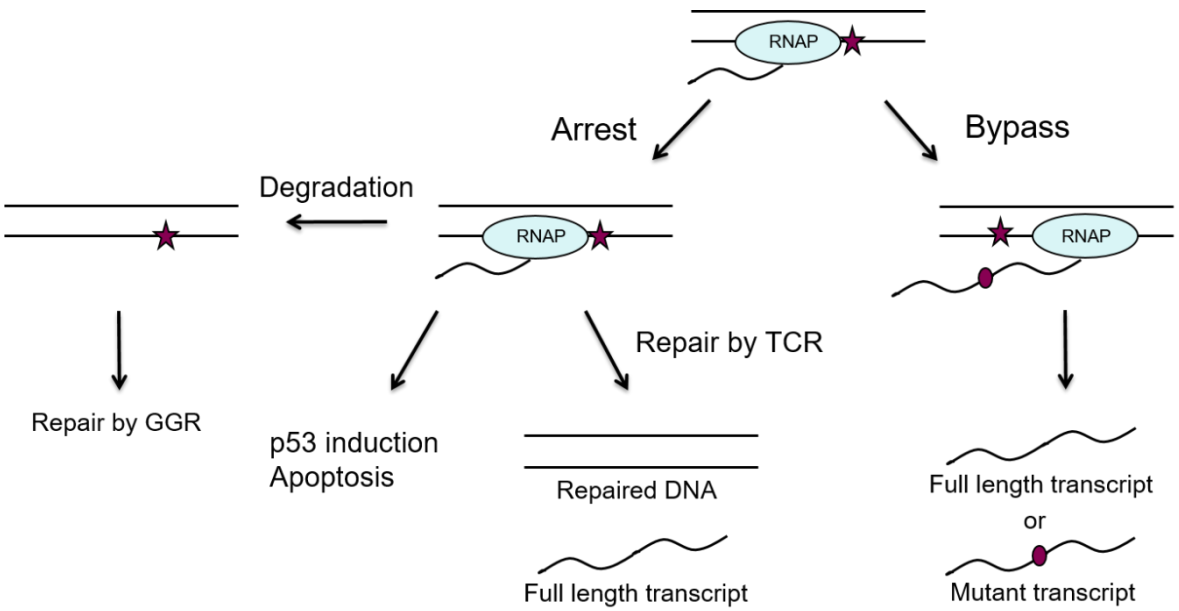

Figure 3. Schematic representation of alternative outcomes of encounters between elongating RNA Pol II and DNA lesion.

DNA lesions can have profound effects not only on RNA Pol fidelity, but also on the survival of cells. Therefore, to protect against deleterious consequences, cells have evolved a specific repair mechanism connected to transcription. RNA Pols II, which are arrested by DNA lesions, initially recruit proteins of the transcription-coupled repair (TC-NER) pathway (see *Section 1.3.1* for details), a specialized nucleotide-excision repair (NER) pathway, that removes the hindering lesion to ensure continuation of transcription (Mellon et al., 1987; Tornaletti and Hanawalt, 1999). Alternatively, in the absence of sufficient rescue, the

irreversibly obstructed RNA Pol can be polyubiquitinated resulting in a rapid degradation by the proteasome through a process viewed as a "last resort" mechanism (Svejstrup, 2002; Svejstrup, 2007). Thus, the gene is freed for transcription by the next RNA Pol II and allows global genomic repair (GG-NER) to recognize and remove the lesion (Figure 3). Mechanisms and consequences of dissociation (prokaryotes) or degradation (eukaryotes) of irreversibly arrested RNA Pol are not within the scope of the thesis and are thus not covered in this review (reviewed in (Svejstrup, 2007)).

The role of RNA Pol II in DNA repair is well-established and allows for the preferential removal of DNA lesions in the template strand of active genes (Mellon et al., 1987; Svejstrup, 2010). Thus, RNA Pol II serves as a specific signal not only for DNA repair, but also for sensing DNA damage in general to protect the cells and maintain the integrity of genome (Derheimer et al., 2007; Ljungman and Lane, 2004). Indeed, RNA Pols have been proposed as the most specific DNA lesion recognition protein (Lindsey-Boltz and Sancar, 2007). In contrast to the well-established responses by stalled RNA Pol, far less is known about consequences of DNA lesion bypass by an elongating RNA Pol. A few studies have shown that different types of DNA lesions can be bypassed with various degrees of efficiency by the elongating RNA Pol II resulting in various extents of misincorporation that generates mutant transcripts via a process termed transcriptional mutagenesis (TM) (Brégeon and Doetsch, 2011; Doetsch, 2002). Thus, TM can lead to the production of mutant proteins that could affect cellular homeostasis in many ways (see *Section 1.4* for details). In summary, in addition to spontaneous errors made by RNA Pol itself, RNA Pol fidelity is greatly influenced by the presence of DNA lesions. The impact of DNA lesions on transcription fidelity could be even more severe and could play an important role in slowly growing or non-dividing cells, in which global DNA repair is greatly diminished or absent (*e.g.* in terminally differentiated mammalian cells) (Dreij et al., 2010; Nouspikel and Hanawalt, 2002).

1.3.1 Transcription-coupled repair

As described above, an encounter of elongating RNA Pol II with a DNA lesion can have severe consequences for the cells as the irreversibly blocked RNA Pol II can induce p53 accumulation and subsequently apoptosis (Derheimer et al., 2007; Ljungman and Lane, 2004). Thus, the removal of DNA lesions from actively transcribed genes is a matter of high priority. Indeed, it was demonstrated some 30 years ago that DNA lesions on the template strand of an actively transcribed gene was preferentially repaired compared to the non-transcribed strand of the same gene or non-transcribed regions of the genome indicating involvement of RNA Pol II in DNA repair (Mellon et al., 1987). The latter DNA lesions are independently recognized and removed by global genome repair pathway (GG-NER) (Hanawalt, 2002). Transcription-coupled repair (TC-NER) has evolved preferentially to target lesions that impede elongating RNA Pol II and was originally documented for UV-induced cyclobutane pyrimidine dimers (CPD) (Mellon et al., 1987). The importance of TC-NER is strengthened by the fact that individuals with mutated TC-NER genes have defects in

repair of lesions in actively transcribed genes and suffer from a severe developmental or neurological hereditary disorder Cockayne's syndrome (CS) (Rapin et al., 2000) and a milder disorder UV sensitive syndrome (UVsS), primarily characterized by skin photosensitivity (Itoh et al., 1994).

Both GG-NER and TC-NER are subpathways of NER and share many protein components except for the initial recognition of a DNA lesion (XPC and DDB for GG-NER, and CSA and CSB for TC-NER). In general, NER is one of the most versatile repair systems and removes a variety of structurally unrelated DNA lesions with an apparent preference for helix-distorting lesions (Hanawalt, 2002; Nouspikel, 2009; Svejstrup, 2002). While TC-NER is initiated by irreversibly stalled RNA Pol II, which is an established sensor for DNA damage on the transcribed strand, GG-NER pathway is triggered by sensing DNA helix distortions elicited by DNA lesions rather than direct lesion recognition (Ljungman and Lane, 2004; Yang, 2007). The core mechanism of NER is well understood and constitutes a number of sequential enzymatic steps, including recognition of a lesion in DNA; opening of denaturation bubble, incisions of the damaged strand one on each side of the lesion, removal of the oligonucleotide containing the lesion, gap filling using the complementary DNA strand as template, and finally, ligation.

The transcription-coupled repair proteins, CSB and CSA, recognize arrested RNA Pols II and are responsible for recruiting the basal NER factors, chromatin remodelers as well as proteins required for subsequent repair-dependent DNA synthesis (reviewed in (Lagerwerf et al., 2011; Mullenders, 2015; Spivak, 2016). CSB, a member of the SWI/SNF protein family, has intrinsic ATP-dependent chromatin remodeling activity, which might be required to open chromatin around lesions thereby stimulating repair (Citterio et al., 2000; van den Boom et al., 2002). In addition, an ubiquitin-binding domain (UBD) has been identified at the C-terminal part of CSB, which is essential for its function in TC-NER (Anindya et al., 2010). CSB is a prerequisite factor that recognizes stalled RNA Pol II, recruits NER proteins (TFIIH, RPA, XPA and two structure specific endonucleases XPG and XPF/ERCC1), histone acetyltransferase p300 and the CSA-E3-ubiquitin ligase complex to the stalled RNA Pol II (Fousteri et al., 2006; Lagerwerf et al., 2011). How CSB identifies stalled RNA Pol II is not fully elucidated, but some evidence suggested CSB being associated with the elongating transcription machinery and that prolonged stalling stabilizes the interaction (Beerens et al., 2005; van den Boom et al., 2002).

CSA is a member of the WD-40 repeat family of proteins, a motif known to be involved in protein-protein interactions, and associates with a cullin-4A containing E3-ubiquitin ligase complex (Groisman et al., 2003; Henning et al., 1995). Assembly of the CSA complex is required for the recruitment of additional TC-NER factors, such as XPA binding protein 2 (XAB2), the high-mobility group nucleosome binding domain 1 (HMGN1) protein, and TFIIS to stalled RNA Pol II/CSB complexes (Fousteri et al., 2006). Moreover, CSB has been shown to be degraded in a CSA-dependent manner, establishing a functional link between CSA and CSB proteins (Groisman et al., 2006). In addition, UVSSA (causative gene for

UVsS) and its partner USP7 have recently been identified having essential roles in regulation of TC-NER, specifically in response to UV induced DNA damage, by stabilizing CSB and facilitating CSA- and CSB-dependent ubiquitination of stalled RNA Poll II at the DNA lesion site (Fei and Chen, 2012; Nakazawa et al., 2012; Schwertman et al., 2012).

1.4 TRANSCRIPTIONAL MUTAGENESIS

Transcriptional mutagenesis (TM) is a replication-independent process for generating mutant transcripts and proteins. In this process, an elongating RNA Pol bypass a DNA lesion on the template strand leading to incorrect nucleotide incorporation opposite the lesion and production of mutant transcripts (Deotsch, 2002). Thus, as long as the DNA lesion remains unrepaired, TM has the potential to generate a substantial population of mutant transcripts, which could be amplified by orders of magnitude during translation, leading to the production of a relatively large amount of the mutant protein (Schwanhausser et al., 2011). As observed in this thesis and by others, the error frequency can increase three-to-four orders of magnitude in the presence of DNA lesions depending on the status of DNA repair compared to the transcription error rate of eukaryotic RNA Pol (10^{-5}). The resulting pool of mutant proteins with potentially altered functions could induce major changes in the cellular homeostasis with a number of deleterious consequences, especially in non-proliferating cells as the capacities of certain DNA repair pathways in such cells are attenuated (Brégeon and Doetsch, 2011; Nouspikel and Hanawalt, 2002). Thus, TM has been proposed to play an important role in human disease development, including tumorigenesis and neurodegeneration (Basu et al., 2015; Brégeon and Doetsch, 2011; Morreall et al., 2013).

TM was discovered almost 30 years ago when researchers investigated how small DNA lesions which do not activate TC-NER affected the transcription machinery. Purified bacteriophage and/or prokaryotic RNA Pols and DNA templates containing single site-specific DNA lesions were used to determine the effects of DNA lesions on transcription elongation *in vitro*. These early studies demonstrated that RNA Pol can bypass DNA lesions (similar to DNA Pol) to different extent and that the bypass can be mutagenic (Liu et al., 1995; Zhou and Doetsch, 1993). Currently, a number of studies have assessed the impact of various DNA lesions on elongating RNA Pol revealing that the extent of TM can vary and depends on different factors. Frequently occurring non-bulky DNA lesions, which induce only small changes to the bases and their base-pairing properties within the DNA, do not block RNA Pol, but are efficiently bypassed to varying extent both *in vitro* and *in vivo* (summarized in (Dreij et al., 2010). These lesions are formed by spontaneous deamination, depurination, alkylation, and oxidation of the bases in DNA and include *e.g.* O^6-methylguanine (O^6-meG) (Burns et al., 2010; Dimitri et al., 2008; Ezerskyte et al., 2018; Viswanathan and Doetsch, 1998), 8-oxoguanine (8-oxoG) (Bregeon et al., 2003; Paredes et al., 2017; Saxowsky et al., 2008), uracil and abasic sites (Bregeon et al., 2003; Kuraoka et al., 2003; Liu et al., 1995; Zhou and Doetsch, 1993).

The mechanism of TM has been investigated using various approaches, such as *in vitro* transcription using DNA templates, structural function analysis or *in silico* modeling, and

revealed that the size and conformation of the DNA lesion are important factors influencing the extent of misincorporation (Damsma and Cramer, 2009; Dimitri et al., 2008; Dreij et al., 2010; Walmacq et al., 2015). In addition, TM can be strongly modulated by sequence context, such as strength of the promoter driving the transcription, sequence flanking the lesion and/or distance from the promoter (Bregeon et al., 2009; Pastoriza-Gallego et al., 2007). For example, *in vitro* studies showed that O^6-meG partially blocks human RNA Pol II elongation, and when the lesion was bypassed (40% of total) and full-length transcripts were obtained, cytidine and uridine were incorporated opposite the lesion at a 3:1 ratio (Dimitri et al., 2008). *In silico* modeling showed that adaptation of an *anti*-conformation around the O^6-meG glycosidic bond with the methyl group in a proximal orientation allowed for transcriptional bypass (Dimitri et al., 2008). Moreover, in the study presented in Paper II, we found that 14.7% of the transcripts contained uridine opposite the lesion in human cells with inactive AGT (Ezerskyte et al., 2018).

In case of 8-oxoG, *in vitro* studies showed no blockage of the elongating RNA Pol II with cytidine insertions opposite to this lesion being favored, although adenosine misincorporations (8%) were also detected (Charlet-Berguerand et al., 2006). A study in murine cells demonstrated that 14% of the transcripts contained a misincorporated adenosine opposite the 8-oxoG (Saxowsky et al., 2008). In addition, structural analysis of 8-oxoG revealed that misincorporated adenosine forms a Hoogsteen base pair with 8-oxoG, which requires an uncommon *syn*-conformation (Damsma and Cramer, 2009). In contrast, bulky lesions, such as cyclobutane pyrimidine dimers, were shown to arrest RNA Pol, thus initiating TC-NER *in vivo* (Selby et al., 1997). However, it was reported that even bulky lesions, such as thymine dimers and 8,5'-cyclo-2'-deoxyadenosine, could be bypassed by human RNA Pol II *in vivo* in mammalian cells, although at low frequency resulting in the production of mutant transcripts (Marietta and Brooks, 2007; Nagel et al., 2014). New studies continuously emerge investigating the potential of different DNA lesions to perturb transcription *in vitro* and in human cells (Xu et al., 2017; You et al., 2015).

A plethora of studies have focused on the mutagenic potential and consequences of DNA lesions during DNA replication to genomic instability and a variety of diseases, including cancer (Hanahan and Weinberg, 2011). In contrast, still very little is known about the impact of TM and these seemingly transient changes in cellular homeostasis. Early *in vivo* studies in bacterial systems using *E. coli* demonstrated that various DNA lesions could be bypassed leading to misincorporation and production of mutant proteins based on luciferase reporters (Bregeon et al., 2003; You et al., 2000). Thus, TM was proposed to enable bacteria to escape from growth restricted environments, and/or to acquire antibiotic resistance (Clauson et al., 2010). Later, using reporter-based systems with site-specific lesions, TM was shown to occur in mammalian cells resulting in the production of proteins with altered functions and suggesting changes in cellular homeostasis due to TM also in higher organisms (Bregeon et al., 2009; Burns et al., 2010). Based on these studies, it was suggested that TM could have detrimental impact on human health by contributing to the etiology of human diseases, including tumorigenesis, aging and neurodegeneration (Brégeon and Doetsch, 2011). Indeed,

accumulating evidence strongly suggests that TM could be a mechanism contributing to tumorigenesis by inactivating tumor suppressors or activating oncogenes as supported by two studies to date (Ezerskyte et al., 2018; Saxowsky et al., 2008). We demonstrated (Paper II and III) that 14.7% of mutant transcripts containing uridine due to O^6-meG-induced TM resulted in the production of mutant p53 (R248W) protein which was sufficient to attenuate its tumor suppressor function (Ezerskyte et al., 2018). In addition, Saxowsky *et al*., demonstrated that transcripts containing adenosine due to TM induced by 8-oxoG resulted in the production of constitutively active RAS (Q61K) protein with subsequent activation of downstream MAPK signaling (Saxowsky et al., 2008). Interestingly, a recent study developed high-throughput assays to investigate individual differences in DNA repair capacity and detected different levels of TM induced by O^6-meG in cells of same origin but from different individuals without known DNA repair deficiencies (Nagel et al., 2014). Moreover, we have shown that TM might affect translation outcome not only by changing the coding specificity of the mutated codon, but also by changing regulatory sequence at a splicing site, thus affecting splicing fidelity (Paredes et al., 2017).

1.5 INTRODUCTION TO THE STUDY

In this thesis, the impact of O^6-meG and 8-oxoG-induced TM on mRNA splicing fidelity and protein function and the consequent effects on regulation of cellular homeostasis has been investigated. In order to determine levels and biologically relevant consequences of TM in mammalian cells, tumor suppressor p53, lamin A and proteolipid protein 1 were used as gene and/or protein probes. The methodology, types of DNA lesion and probes are described and discussed below.

1.5.1 How to study transcriptional mutagenesis

The first studies investigating TM were performed using *in vitro* transcription systems comprised of purified components and/or cellular extracts from various types of cells sufficient to initiate and push forward transcription reaction on a DNA template (Chen and Bogenhagen, 1993; Kuraoka et al., 2003; Viswanathan and Doetsch, 1998). A summary of various *in vitro* and *in vivo* studies addressing distinct facets of TM induced by a number of different DNA lesions is provided in review article by Brégeon and Doetsch (Brégeon and Doetsch, 2011). RNA Pols from bacteriophages, bacteria or mammals were employed in these *in vitro* studies to assess whether various lesions allow bypass of RNA Pol and whether bypass results in misincorporation producing a mutated product. Linear DNA templates for transcription contained single, well-characterized DNA lesions at specific positions. *In vitro* studies can be a useful and fast tool to quickly examine if particular DNA lesions exert mutagenic or blocking effects during transcription.

In order to address the same questions and what the subsequent biological consequences of TM are *in vivo*, the expression constructs need to contain well-characterized DNA lesions at specific positions in genes/proteins of interest. Such experimental systems have to meet two requirements. The first requirement is to design a plasmid encoding a reporter protein whose

activity could be easily monitored biochemically, *e.g.* luciferase or red fluorescent protein (RFP) (Burns et al., 2010; Shaw et al., 2002) or functionally, *e.g.* RAS or p53 tumor suppressor (Ezerskyte et al., 2018; Saxowsky et al., 2008). Secondly, to ensure that the measured activity of the reporter protein is induced by TM and not due to a mutation that is fixed within the sequence after DNA replication. To eliminate the latter problem in bacterial systems, cells can be maintained in non-growing conditions, in which only transcription occurs, by the use of antibiotics, whereas, in eukaryotic systems, plasmids without known origins of replication should be used (Brégeon and Doetsch, 2011).

Several protocols with different advantages and limitations have been developed to produce sufficient quantities of plasmid to study the effects of an individual lesion *in vivo* (Bregeon and Doetsch, 2004; Burns et al., 2010; Luhnsdorf et al., 2012; You et al., 2012; You et al., 2000). Two different well-established strategies, "gapped duplex" and "primer extension", were used in this thesis to produce reporter plasmids to study effects of DNA lesion-induced TM in mammalian cells (see *Section 3.1* for detailed description).

In addition to transcription reporter plasmids containing site-specific DNA lesions in target genes, advances of next-generation RNA sequencing technology provides another approach to determine multiple facets of TM at a genome-wide level. Several recent studies employing RNA-seq have already provided with insights about the spectrum of spontaneous transcriptional errors due to RNA Pol infidelity and effects of these errors to proteostasis and subsequently to cellular homeostasis (Gout et al., 2017).

1.5.2 DNA damages and their repair

1.5.2.1 O^6-methylguanine

Even though O^6-methylguanine (O^6-meG) is not the most abundant DNA lesion, it is a highly mutagenic lesion induced by both endogenous (*e.g. S*-adenosyl methionine [SAM]) and exogenous (*e.g. N*-nitroso-*N*-methylurea [MNU], *N*-methyl-*N′*-nitro-*N*-nitrosoguanidine [MNNG], chemotherapeutics temozolomide) alkylating agents (De Bont and van Larebeke, 2004; Hemminki, 1983; Rydberg and Lindahl, 1982; Tisdale, 1987). O^6-meG can cause base mispairing and thus instructs DNA Pol to incorporate thymidine instead of cytidine opposite the lesion, resulting in GC → AT transitions during replication (Abbott and Saffhill, 1979; Singh et al., 1996). In addition, it has been shown that O^6-meG can be bypassed by RNA Pol II *in vitro* and *in vivo* directing the insertion of uridine instead of cytidine into a nascent RNA transcript (Burns et al., 2010; Dimitri et al., 2008; Ezerskyte et al., 2018; Viswanathan and Doetsch, 1998). While primarily considered as a mutagenic lesion, O^6-meG is also cytotoxic in certain circumstances. If not repaired by AGT, replication of DNA containing O^6-meG gives rise to O^6-meG:T mispairing, which invokes DNA mismatch repair (MMR) pathway. However, MMR generates a long gap in the newly synthesized strand and synthesis results in regeneration of the O^6-meG:T base pair via process referred to as a futile MMR cycle that ultimately produces double strand DNA breaks and induces apoptosis (Kaina et al., 2007; Margison and Santibanez-Koref, 2002).

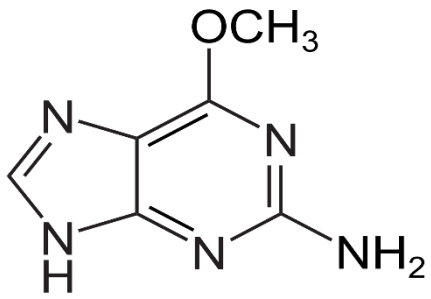

Figure 4. Chemical structure of O^6-methylguanine.

The severe impact of this lesion to the genomic stability is supported by the fact that O^6-meG is directly eliminated in a single-step error-free reaction by the ubiquitous repair protein, O^6-alkylguanine-DNA alkyltransferase (AGT), which has been identified in all studied organisms and is encoded by the *MGMT* gene in humans (Margison and Santibanez-Koref, 2002). This protein is referred to as a suicidal protein and it accepts the alkyl group from the lesion to a cysteine residue in its active site in the single-step stoichiometric reaction (Tano et al., 1990). Following the transfer of the alkyl group, the active site is not regenerated and the inactive enzyme is ubiquitinated and degraded by the proteasome (Margison and Santibanez-Koref, 2002; Mishina et al., 2006). AGT is highly efficient in suppressing point mutations and genotoxicity induced by O^6-meG *in vitro* and *in vivo* and the most important defense against its tumorigenesis (Kaina et al., 2007). However, expression of AGT varies extensively in different human tissues as well as in tumors. Repression of *MGMT* gene expression due to hypermethylation of the *MGMT* gene promoter is the major reason for inactivation of O^6-meG repair in cell lines and tumors, such as brain or lung, rather than a mutation within the gene (Christmann et al., 2011; Halford et al., 2005). In addition, tumors that overexpress *MGMT* are resistant to chemotherapeutic O^6-alkylating agents, such as temozolomide.

1.5.2.2 8-oxoguanine

The most mutagenic and best-characterized lesion formed by reactive oxygen species (ROS) is 8-oxoguanine (8-oxoG) (Lindahl, 1993). 8-oxoG is highly mutagenic as it can assume a *syn*-conformation mimicking thymidine and base pair with adenosine resulting in GC → TA transversion mutation during replication (Cheng et al., 1992; Damsma and Cramer, 2009; Shibutani et al., 1991). Additionally, 8-oxoG can be bypassed by RNA Pol both *in vitro* and *in vivo*, resulting in cytidine or mutagenic adenosine incorporations opposite the lesion, as well as -1 deletions of transcripts (Bregeon et al., 2003; Bregeon et al., 2009; Chen and Bogenhagen, 1993; Damsma and Cramer, 2009; Kuraoka et al., 2003; Paredes et al., 2017; Saxowsky et al., 2008; Viswanathan and Doetsch, 1998).

Various small, non-helix-distorting oxidative DNA damages, such as 8-oxoG, are recognized and removed by DNA glycosylases, which initiate the base excision repair (BER) pathway. DNA glycosylases hydrolyze the N-glycosylic bond forming an apurinic/apyrimidinic (AP) site in DNA which is further processed by other types of enzymes which are part of the BER pathway (Krokan et al., 1997). 8-oxoG is the primary physiological substrate for the 8-

oxoguanine DNA glycosylase (OGG1) (Tchou et al., 1991). Increased accumulation of 8-oxoG in DNA was detected in *OGG1* defective mice compared with wild type *OGG1* mice. Defective mice were also found to have a higher risk for developing spontaneous lung adenoma/carcinoma (Sakumi et al., 2003). In human cancers, BER repair pathway can be inactivated by the downregulation of *OGG1* gene expression (Jiang et al., 2006). A study has discovered that *OGG1* expression was deregulated epigenetically by both hyper- and hypomethylation of promoter regions in different breast tumor samples (Fleischer et al., 2014).

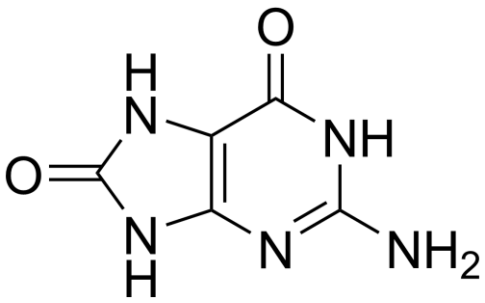

Figure 5. Chemical structure of 8-oxoguanine.

Although TC-NER is activated by DNA damages that pose an irreversible arrest for elongating RNA Pol II, it has also been reported that oxidative damage, such as 8-oxoG, could be subjected to TC-NER in *mfd*-deficient *E. coli* or in *CSB*-deficient mouse embryonic fibroblasts (Bregeon et al., 2003; Pastoriza-Gallego et al., 2007). However, to date evidence is lacking to support this hypothesis, as a number of key articles supporting oxidative damage repair by TC-NER have been retracted (Hanawalt and Spivak, 2008). In addition, a study with cells defective for both TC-NER and the repair of oxidative lesions, revealed that 8-oxoG is not repaired by TC-NER pathway (Bregeon et al., 2009).

1.5.3 Reporter proteins

1.5.3.1 The guardian of the genome - p53 protein

In the thesis, we used one of the most important tumor suppressor proteins as a probe to study cellular effects of TM induced by O^6-meG. The p53 protein, the guardian of the genome, primarily functions as a transcription factor regulating expression of numerous target genes and mediating a variety of cellular responses, such as cell cycle arrest and apoptosis (Horn and Vousden, 2007). Different cellular stresses, such as DNA damage or oncogene activation, stabilize and activate p53. Critical to the tumor suppressor function of p53 is its ability to bind sequence-specific DNA sites, thus making p53 a susceptible target for mutations. Indeed, biological consequences of mutations in the *TP53* gene are evident since mutant p53 is detected in more than half of all human tumors. Moreover, the vast majority of these mutations are missense point mutations with dominant-negative properties abrogating ability of p53 to transactivate its target genes (Bullock and Fersht, 2001; Petitjean et al., 2007).

As a probe to study effects of O^6-meG-induced TM on p53 function, the arginine to tryptophan mutation at codon 248 (c. 742 C>T, R248W) was used. This well-studied hot spot mutation with dominant-negative phenotype is commonly found in several human cancers and thus of high significant biological relevance (Petitjean et al., 2007). We placed a single O^6-meG lesion at a defined position within codon 248 (G*CC on the template strand and CGG coding sequence) of *TP53* gene in an expression construct. Bypass of the lesion by RNA Pol II during transcription could have two possible outcomes. If the lesion directs insertion of cytidine, codon CGG for arginine is produced subsequently resulting in synthesis of wild type p53. However, if RNA Pol II misreads at the site of lesion and inserts uridine, the transcript with the codon UGG for tryptophan is synthesized that subsequently is translated into mutant R248W p53 protein. Thus, events of TM induced by O^6-meG can produce a pool of wild type and mutant p53 within a cell. Two control vectors encoding for wild type and mutant R248W p53 were also constructed for cellular responses to the expression of p53 variants. The expression construct also encoded GFP that served as an internal control for transfection and assisted in the analysis. Studies have shown that mutant R248W p53 abrogates the tumor suppressor function of p53 affecting expression of target genes, which, in turn, results in deregulated cell cycle control (Willis et al., 2004). Thus, p53 is a suitable probe protein to study effects of TM induced by O^6-meG in mammalian cells and to characterize the potential link between TM and cancer etiology.

1.5.3.2 Lamin A and proteolipid protein 1

As discussed above, alternative splicing is a highly regulated process that requires extreme precision. Effects of splicing infidelity can have detrimental biological consequences that are reflected by a large number of severe diseases associated with aberrant splicing (Ward and Cooper, 2010). Indeed, transcriptional errors might affect translation outcome not only by changing the coding specificity of a codon, but also by changing the sequence of a splicing signal that would affect the recognition of the splice site. The ability of RNA Pol errors to significantly affect splicing fidelity has been proposed before (Doetsch, 2002; Fox-Walsh and Hertel, 2009), and recently suggested from data obtained measuring RNA Pol fidelity using RNA sequencing data (Carey, 2015). However, the impact of lesion-induced TM on splicing fidelity has not yet been investigated. To study the impact of TM on splicing fidelity we chose two relevant human genes, *lamin A* (*LMNA*) and *proteolipid protein 1* (*PLP1*), with established links between aberrant splicing and disease (Eriksson et al., 2003; Hobson et al., 2006).

Lamin A proteins, encoded by the *LMNA* gene in humans, also known as intermediate filaments, are structural proteins that provide a scaffold for the cell nucleus and form the nuclear lamina, which plays crucial roles in cell division, DNA replication, repair, gene transcription and chromatin remodeling and are thus very import regulators of cellular functions (Davies et al., 2009). Mutations in *LMNA* gene or defective posttranslational processing cause the majority of human genetic diseases termed laminopathies. Missense mutations are a major reason for diseases. Hutchinson-Gilford progeria syndrome (HGPS),

characterized by premature aging, is caused by a mutant prelamin A that cannot be processed to a mature lamin A (Eriksson et al., 2003). HGPS individuals most commonly carries *de novo* heterozygous C to T base substitution (c.1824C>T) in codon 608 leading to a silent G608G mutation within exon 11 (CAG′GTGGGT). This mutation increases the usage of a cryptic splice site resulting in the removal of additional 150 nucleotides, which cause a deletion of 50 aa near the C-terminus of lamin A producing a protein called progerin (Eriksson et al., 2003). We utilized this mutation as a probe in this thesis to study effects of O^6meG on splicing fidelity in mammalian cells. We placed O^6-meG in codon 608 on the template strand (CCG*, coding sequence GGC). During transcription, transcripts containing either the wild type GGC codon or the mutated GGU codon due to uridine misincorporation opposite O^6-meG, causing the same splicing defect as in HGPS, should be produced.

Proteolipid protein 1 (PLP1) is the major myelin protein of the central nervous system (CNS) (Woodward and Malcolm, 1999). PLP1 and its splicing variant DM20, both simultaneously expressed, are transmembrane proteins playing an important role in myelination process. Both forms are generated from the same primary transcript by joining two competing 5' splice sites in exon 3 to the same 3' splice site by alternative splicing (Hobson et al., 2006). Changes in the dosage between the two splice variants and mutations in *PLP1* can cause the Pelizaeus–Merzbacher disease (PMD) (Woodward and Malcolm, 1999). To study the impact of 8-oxoG-induced TM on splicing fidelity, we employed the threonine to lysine mutation at codon 116 (T116K) within exon 3 (AAG′GTAACA) that is caused by a single C to A base substitution (c.347C>A). This mutation promotes alternative placing of DM20 which is associated with PMD (Nance et al., 1996). The 8-oxoG was placed in codon 116 on the template strand (TG*C, coding sequence ACG), which during transcription could produce transcripts containing either the wild type codon ACG encoding threonine or the mutated AAG codon due to adenosine misincorporation directed by 8-oxoG.

2 AIM OF THE STUDY

Despite numerous mechanistic studies revealing that various DNA lesions are efficiently bypassed by elongating RNA Pol, biologically relevant consequences of TM that could affect cellular homeostasis remain to be fully elucidated. The overall aim of this thesis was to characterize the occurrence and to shed some light on the impact of TM on cellular homeostasis that could possibly play a role in disease and tumor development.

The specific aims were as follows:

- To develop a GFP-based minigene splicing reporter system containing a site-specific O^6-meG and 8-oxoG (Paper I)
- To construct a plasmid based on the *TP53* and *GFP* genes with a site-specific O^6-meG (Paper II)
- To investigate the impact of lesion-induced TM on splicing fidelity in mammalian cells (Paper I)
- To examine the impact of O^6-meG-induced TM on p53 function as a tumor suppressor in human cells (Paper II)
- To investigate the impact of O^6-meG-induced TM on the p53 transactivation ability at a genome-wide level (Paper III)

3 METHODS AND CONSIDERATIONS

This section provides a brief overview of methods used in the constituent publications of this thesis with the focus on their advantages as well as limitations. Detailed protocols of the techniques and methods are presented in the Materials and Methods of each constituent article.

3.1 CONSTRUCTION OF DAMAGED PLASMIDS

Non-replicative transcription templates containing a single DNA lesion at a defined position in the transcribed strand of a reporter gene are required to study biologically relevant effects of mutated transcripts generated through TM. For the study in Paper I, we produced two minigene splicing reporter systems using *LMNA* and *PLP1* to study effects of TM induced by O^6-meG and 8-oxoG, respectively, on splicing fidelity. In Paper II and III, in order to study effects of TM induced by O^6-meG on protein function, we constructed plasmids using *TP53* as a reporter. Each reporter expression system contained three plasmids that encoded the wild-type sequence of the reporter gene, the mutated sequence with specific point mutation, or the wild-type sequence with a DNA lesion positioned within a codon of interest, thus mutagenic bypass of the lesion would result in the production of mutant protein. Development of control plasmids containing wild-type or mutant reporter sequences will not be described here as detailed protocols can be found in the constituent publications. The well-established techniques, "primer extension" and "gapped duplex", which were used for construction of the site-specifically damaged plasmids are discussed below. Detailed description of reagents and procedures are provided in appended publications. Both approaches provided a versatile tool using simple standard DNA manipulation techniques to alter a defined site within a region of a plasmid that could be used in many applications. However, both methods are time-consuming and challenging multistep procedures. Conversion of ssDNA to modified dsDNA is inefficient and always incomplete and requires further purification steps, such as cesium chloride density gradient centrifugation.

3.1.1 Minigene reporter constructs

Minigene splicing reporter plasmids containing site-specific DNA damage for *LMNA* and *PLP1* were constructed using the well-established "primer extension" approach as described previously (Bregeon and Doetsch, 2004; Zoller and Smith, 1983). The main steps of "primer extension" method are summarized in Figure 6. The principle of the method is the extension of an oligonucleotide primer annealed to a single stranded template by DNA Pol. Briefly, modified oligos (O^6-meG oligo for LMNA and 8-oxoG oligo for PLP1) in molar excess were annealed to purified wild-type ssDNA, which was propagated in bacteria infected with helper phage, and extension of the complementary strand was performed. Then closed circular dsDNA (cc dsDNA) was formed by ligation. Different steps of this protocol were optimized by Bregeon and Doetsch to improve yields of site-specifically modified cc dsDNA, that micrograms of the final product can be produced (Bregeon and Doetsch, 2004). However, this method has several drawbacks. First, there is a risk, especially with larger plasmids, that

the damaged oligo could anneal unspecifically to a similar sequence at another region of the plasmid. Additionally, the size of plasmid could also be a limiting factor for a successful second strand synthesis and even the final ligation reaction might be sequence context dependent and impossible to achieve. Moreover, purified DNA Pol has a high error frequency, thus additional mutations outside the modified oligo may be expected. However, this approach requires relatively small amounts of starter material as few purification steps are performed.

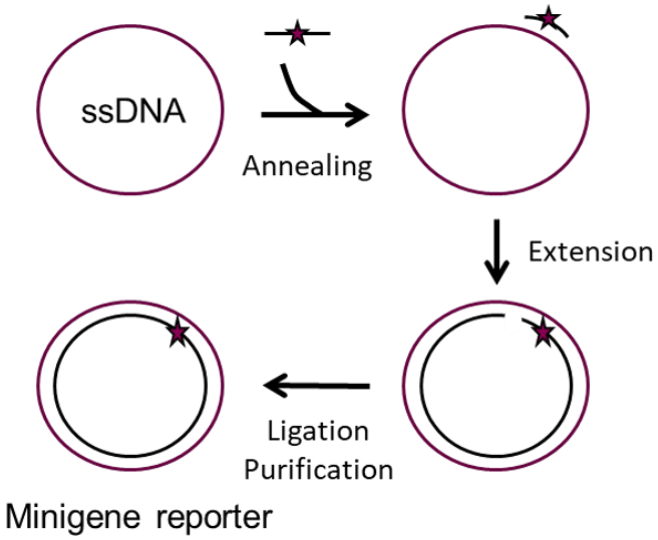

Figure 6. Schematic presentation of the major steps of the "primer extension" method to produce a site-specific lesion containing minigene splicing reporter plasmid.

Worth noting, the protein specific exon-intron-exon sequences for each probe protein were used in combination with an EGFP open reading frame in our study. Alternative splicing is activated for each gene when specific mutations (C>T or C>A) occur in their splicing signal sequences as described above (see *1.5.3.2*). In order to detect the activation of alternative splicing due to TM, the minigene reporters were constructed in a way that EGFP is only expressed if alternative splicing has occurred. This was achieved by introducing a STOP codon at the end of the first exon containing the alternative donor splice site (Figure 7). When alternative splicing occurs due to misincorporation event of U (induced by O^6-meG) or A (induced by 8-oxoG) into pre-mRNA during transcription, the STOP codon is removed together with adjoining intron. As a result, GFP is expressed and can be detected to evaluate activation of alternative splicing. To compare cellular responses to expression of different protein variants, two control vectors containing either wild-type or corresponding mutated sequences for both reporters were constructed.

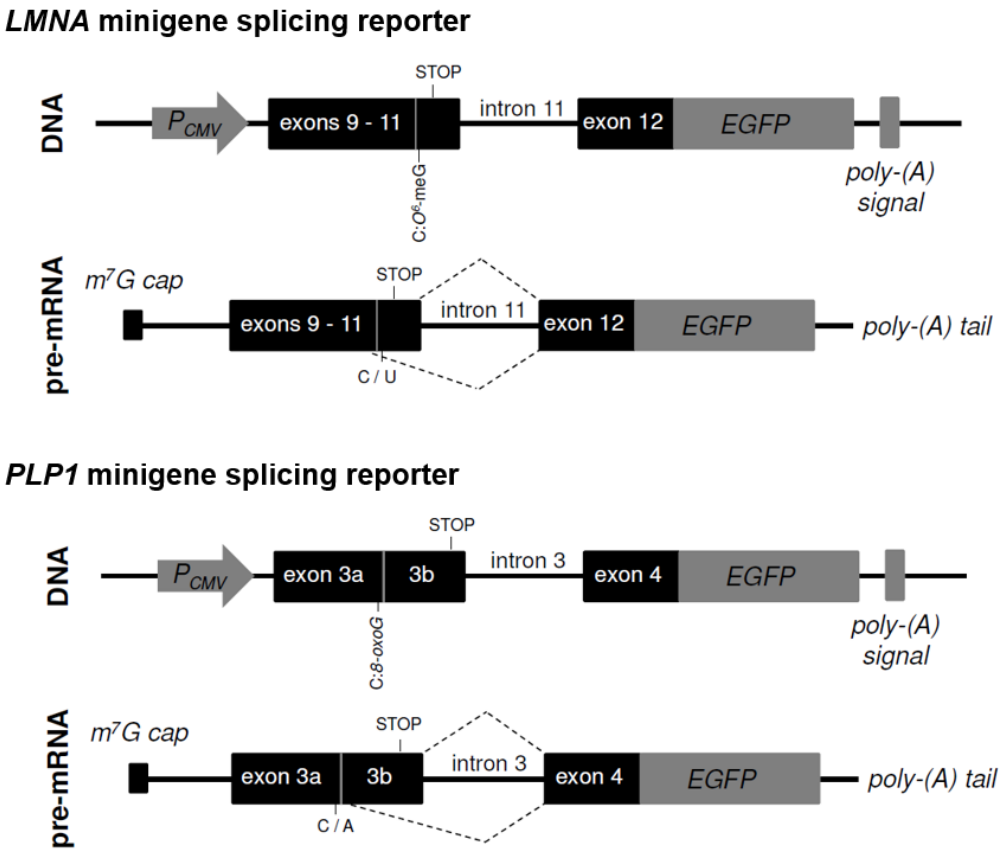

Figure 7. Schematic presentation of the minigene splicing reporters containing site-specific DNA lesions with their resulting pre-mRNA used in this study. Image reused from Paredes et al., 2017 with modifications.

3.1.2 p53 TM plasmid

Synthesis of a p53 plasmid containing a site-specific O^6-meG was accomplished using the "gapped duplex" method as described previously (Burns et al., 2010; Kramer and Fritz, 1987; Tornaletti et al., 1997) with some modifications. A summary of main steps of the "gapped duplex" protocol is presented in Figure 8. The principle of the method is the formation of ssDNA/DNA heteroduplex that contains a gap by use of the oscillating phenol emulsion reassociation technique (OsPERT) as previously described (Bruzel and Cheung, 2006). In our system, the gapped duplex contained an 11 nucleotide gap to which an 11-bp oligomer with a site-specific damage was annealed and ligated to form cc dsDNA. With this procedure up to 100 µg pure cc dsDNA could be produced. However, this approach has some drawbacks when compared to the above discussed "primer extension" approach. Most importantly, this method requires large amounts of starting material (milligrams of ssDNA) as the annealing process to form gapped duplex DNA is highly inefficient and additional purification steps are required. However, annealing of the damaged oligo is targeted and very specific compared to the other method and helps to minimize unwanted side reactions.

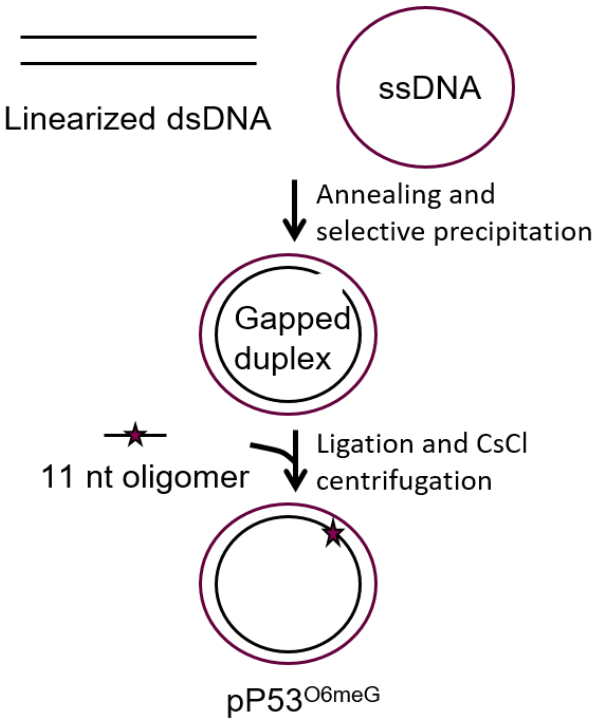

Figure 8. Schematic presentation of the major steps of "gapped duplex" method to produce a site-specific lesion containing plasmid.

3.2 CELL LINES

We used WI-38 fibroblasts that are a normal diploid human cell line derived from lung tissue of 3 months gestation female fetus (Hayflick, 1965). WI-38 cells have a finite number of 50 (±10) population doublings and are not tumorigenic. Fibroblasts were purchased from ATCC and maintained in MEM supplemented with 1 mM sodium pyruvate, 10% FBS, and antibiotics according to the provider's guidelines. WI-38 fibroblasts were selected to study impact of O^6-meG-induced TM on splicing fidelity (Paper I), as these non-tumorigenic cells with normal cellular morphology are suitable for transfections yielding acceptable efficiencies. Furthermore, it is a biologically relevant model reflecting to some extent *in vivo* environment as lungs are one of human organs that can be heavily affected by air-borne alkylating agents from tobacco smoke or environmental pollution.

Mouse embryonic fibroblasts, MEFs, both wild-type *OGG1* and *OGG1*$^{-/-}$ were obtained from Dr Pablo Radicella, The French Alternative Energies and Atomic Energy Commission (CEA), France. These MEFs were originally generated from mouse embryos 13.5 days (E13.5) after gestation by Prof Arne Klungland (Klungland et al., 1999). They are characterized by a finite lifespan of several population doublings and very limited expansion capacity. MEFs were maintained according to the provider's guidelines in DMEM/F:12 (3:1) supplemented with 10% FBS. 2 mM L-glutamine, 1 mM sodium pyruvate and antibiotics. MEFs were used to determine levels of 8-oxoG-induced TM and to study impact of TM on splicing fidelity in mammalian cells (Paper I). To study the importance of DNA repair in preventing TM, the activity of OGG1 enzyme has to be impaired using either an OGG1 inhibitor or *OGG1* deficient cells. The MEF model was chosen for its particular suitability to

the scope of the study, even though they are murine cells. Potent and selective OGG1 inhibitors are still under development and, to the best of our knowledge, no *OGG1*-null cells of human origin have been generated.

The human embryonic kidney cell line, HEK293, was obtained from Dr Katarina Johansson, Karolinska Institutet. HEK293 cells were generated after immortalization of normal human embryonic kidney cells by adenovirus type 5 sheared DNA (Graham et al., 1977). These cells are described as hypotriploid with the modal chromosome number 64 occurring in 30% of cells. Furthermore, HEK293 is tumorigenic cell line forming tumors in nude mice, however with low efficiency (Graham et al., 1977). HEK293 cells were cultured in MEM supplemented with 1 mM sodium pyruvate, non-essential amino acids, 10% FBS, and antibiotics according to the manufacturer's guidelines. This cell line was chosen in Paper I for its fast growth and high transfection efficiencies to determine levels of TM induced by O^6-meG and 8-oxoG and to investigate the impact of TM on splicing fidelity in human-derived cells as a reference cell line to WI-38 and MEFs.

The H1299 human non-small cell lung carcinoma cell line was purchased from ATCC and used in Paper II and Paper III. H1299 were derived from a lymph node metastasis of the lung from a patient who had received prior radiation therapy (Giaccone et al., 1992). These cells possess a capacity to divide indefinitely. H1299 cells were grown according to the manufacturer's instructions in DMEM supplemented with 10% FBS, 1 mM L-glutamine, and antibiotics. These cells are *TP53*-null and do not express p53 protein. The H1299 model was chosen to study consequences of O^6-meG-induced TM on p53 function as a tumor suppressor to avoid any interference from endogenous p53 when interpreting results.

3.3 TRANSFECTION

Transfection is the process of artificial delivery of nucleic acids (DNA or RNA) into eukaryotic cells by physical (*e.g.* electroporation), chemical (*e.g.* cationic lipids) or biological (*e.g.*viruses) approaches. Each transfection method has advantages and drawbacks, but the most important criteria for choosing are an optimal balance between high transfection efficiency, low cytotoxicity, versatility, and convenience. Various factors, such cell type or plasmid size, are known to influence transfection efficiencies. For this reason, trial experiments using various methods and reagents were performed to find optimal transfection method for each cell type leading to satisfying balance between transfection efficiency and cytotoxicity.

To study the effects of TM in mammalian cells, the constructed reporter plasmids were introduced into cells by transfection. In Paper I, transient transfections of WI-38 fibroblasts and MEFs delivering *LMNA* and *PLP1* minigene splicing reporters, respectively, were accomplished using the Amaxa Nucleofector 2B (Lonza) electroporation-based system in accordance with the manufacturer's protocols. Nucleofection is a physical electroporation-based transfection method that uses a combination of specific voltage parameters generated by a Nucleofector device and cell-type specific reagents to transfer substrate into the cell

nucleus and the cytoplasm. High-voltage electrical pulses applied for a short time perturb the plasma membrane creating temporary pores and allowing nucleic acids to enter the cells (Neumann et al., 1982). It is a rapid technique allowing transfection of a large number of all cell types in a short time, but can be tedious to determine optimal conditions and can cause substantial cytotoxicity.

In Paper I, HEK293 cells were transfected with all minigene splicing reporter plasmids using Lipofectamine 3000 (Invitrogen) following manufacturer's instructions. In Paper II and III, delivery of p53 TM plasmids into H1299 cells was performed using Lipofectamine 2000 (Invitrogen) according to manufacturer's guidelines. Cationic lipids, such as Lipofectamine 2000 or 3000, facilitate DNA delivery to the cells by binding spontaneously to the negatively charged DNA and forming complexes that are attracted to the cell membrane and internalized (Pack et al., 2005). This method is easy to perform and yields in relatively high transfection efficiencies; however, it works only with limited number of cell types and can be cytotoxic.

3.4 FLOW CYTOMETRY

Flow cytometry is a powerful laser-based technology to analyze multiple characteristics of a single cell, such as size and granularity, as the cell travels in suspension. Scattered lights at several angles and fluorescence emissions from thousands of cells per second colliding with a laser beam are detected and the signal is transformed into statistical data that can be presented in various graphic formats. Forward scatter (FSc) is collected along the same axis as the laser beam and gives information about cell size, whereas internal complexity and granularity of the cell is represented by side scatter (SSc) (Bakke, 2001). Flow cytometry is a widely used method with a numerous applications, such as analysis of cell size, characterization of different cell types in a heterogeneous cell population, immunophenotyping, cell isolation, assessing the purity of isolated subpopulations, cell cycle analysis, and proliferation assay. The two greatest advantages of flow cytometry are its ability to analyze thousands of cells in a matter of seconds while measuring a large number of parameters on the same sample. The major limitations of flow cytometry are difficulties to differentiate cell subpopulations that express similar markers and loss of information on tissue architecture and cell-cell interactions due to the requirement of a single cell suspension for analysis. Moreover, use of multiple fluorophores can be limited due to fluorescence spillover, which is sometimes impossible to compensate.

Transfection efficiencies of the plasmids were measured by flow cytometry using BD Accuri C6. For the p53 plasmids, we used the co-expressed GFP as a marker (Paper II and III). To determine transfection efficiencies of minigene splicing reporter plasmids that contain modified GFP gene, simultaneous co-transfections with a plasmid encoding mCherry, a monomeric red fluorescent protein, were performed (Paper I). Equimolar concentrations of plasmids were used in co-transfections. In addition, flow cytometry was used to determine whether O^6-meG- and 8-oxoG-induced TM influences splicing fidelity in mammalian cells transfected with minigene splicing reporter plasmids. The GFP was expressed in the transfected cells only if TM occurred activating an alternative splicing site. Changes in GFP

fluorescence intensities after gating GFP-positive cells relative to untransfected cells in response to different conditions were measured with BD Accuri C6.

Fluorescence-activated cell sorting (FACS) was used to purify transfected cells based on the fluorescence signal of the GFP reporter 6h and 24h after transfection (Paper II and III). Cell sorting was performed at the Flow Cytometry Core Facility, Karolinska institutet using MoFlo XPD, BD FACSVantage Diva, BD FACSAria™ III, and BD FACSAria™ Fusion.

Furthermore, we used flow cytometry to perform a bivariate cell cycle analysis of sorted GFP-positive cells (Paper II) in order to examine the effect of p53 TM on cell-cycle control. Cell cycle analysis using PI to quantify DNA content was one of the earliest application of flow cytometry (Krishan, 1975) and remains a widely used application of flow cytometry. To investigate the impact of p53 TM on induction of apoptosis (Paper II), mitochondrial membrane potential (MMP) in live cells was measured using tetramethylrhodamine ethyl ester (TMRE) staining with BD Accuri C6. The signal of GFP expression was used to gate for transfected cells following discrimination of dead cells and doublets.

3.5 QUANTIFICATION OF TM

Two different methods were applied in order to determine levels of TM in mammalian cells in this thesis. We employed Sanger sequencing of RT-PCR products with or without additional PCR-based pre-screening of mutated transcripts and RNA sequencing. Both methods are valid for determination of TM levels and the choice depends on the scope and aims of the study. The screening of cellular transcripts by sequencing of PCR products is a cost-effective but laborious method that can be performed with standard laboratory equipment, whereas RNA sequencing is a rapid but relatively expensive method, which requires special sequencing instruments and expertise in bioinformatics to analyze obtained data. On the other hand, RNA-seq is a high-throughput approach, which enables analysis of thousands of transcripts and provides information about the whole transcriptome at a single nucleotide level in a single run, which cannot be obtained with other approaches.

Sequencing of RT-PCR products was used to determine levels of TM induced by both site-specific O^6-meG and 8-oxoG in *LMNA* and *PLP1* minigene reporters, respectively, in mammalian cells. For *LMNA* reporter, transformed bacteria colonies with cloned RT-PCR products of cellular transcripts were directly subjected to Sanger sequencing, whereas pre-screening step was performed for *PLP1* reporter utilizing tetra-primer ARMS-PCR (Ye et al., 2001). This approach included colony PCR using specially designed set of primers and product resolving on agarose gel to screen for mutated transcripts. Colonies that amplified with mutant-specific primers were further confirmed by Sanger sequencing. In addition, high levels of p53 TM induce by O^6-meG at early time-point in Paper III were confirmed by sequencing of RT-PCR products with pre-screening step of tetra-primer ARMS-PCR to screen for mutated transcripts as described above.

RNA sequencing approach (method described in *Section 3.6*) was used to determine levels of TM induced by site-specific O^6-meG in *TP53* gene in human H1299 cells (Paper II and III).

In addition, *in vitro* transcription assay was performed utilizing DNA templates with a site-specific lesion and RNA pol II from HelaScribe nuclear extract to determine TM levels *in vitro*. Results from *in vitro* transcription assay might not reflect *in vivo* scenario but is one of the ways to show quickly that a particular DNA lesion could induce TM.

3.6 RNA SEQUENCING

RNA-seq is a powerful technique of next-generation sequencing (NGS) that allows the entire transcriptome of any biological sample to be analyzed in a high-throughput and quantitative manner. RNA-seq has versatile application possibilities, as this method does not require knowing the sequence *a priori*, thus facilitating the discovery of novel transcripts. Compared to previous low-throughput technologies, such as microarray- or Sanger sequencing-based approaches, RNA-seq provides high coverage and a single base resolution, very low background noise and large dynamic range (>8000-fold), and is able to distinguish different isoforms or even sequence variations (SNPs) (Kukurba and Montgomery, 2015; Wang et al., 2009). In addition, RNA-seq enables monitoring of gene expression between various samples (*e.g.* disease and normal tissues), conditions or developmental stages and has become an invaluable tool in disease diagnostics as well as research.

A typical RNA-seq experiment includes the following main steps:

- cDNA library preparation
- cluster amplification
- sequencing by synthesis (Illumina)
- alignment and data analysis

Briefly, the first step in cDNA library preparation is the isolation of RNA from a biological sample, which then is reverse transcribed to produce cDNA. Following subsequent second strand synthesis, adaptors specific for each cDNA library and enabling simultaneous sequencing of multiple cDNA libraries are ligated to each end of ds cDNA fragment. Then, the library is enriched by a PCR reaction using adaptor sequences as primers; finally, a quality control is performed followed by normalization and pooling of the libraries if several samples are being sequenced.

Next, for the Illumina sequencing platform that was used in this thesis, library is loaded into a flow cell where the generated fragments hybridize complementary to the surface-bound oligomers. Each fragment is then amplified into distinct, clonal clusters through bridge amplification. The Illumina platform uses sequencing by synthesis approach. Once cluster generation is complete, sequencing starts when primers attach and the first fluorescent labelled terminators, which are version of nucleotides that stop DNA synthesis and each of the four bases have unique emission, are incorporated allowing imaging of the fluorescence signal before another nucleotide addition. The emission wavelength and intensity are used to identify incorporated bases and the cycle continues to create a read length of particular number of bases. The sequencing can be performed single read mode (from one end of the

fragment) or paired-end read mode (from both ends of the fragment). The latter results in more accurate read alignment especially across repetitive regions of sequence, while some approaches are best served by single read mode sequencing (Illumina, 2017). RNA-seq was performed in Paper II and III.

After sequencing, generated read data is then aligned to a reference genome if available or *de novo* assembly is performed. Then the mapped reads can be assembled into transcripts and a quantification of gene expression levels can be performed by counting the number of reads that mapped to full transcripts (Kukurba and Montgomery, 2015). Then, this data can be used to perform a differential gene expression analysis using several available RNA-seq software packages. In this thesis we used the DESeq2 package that uses counts as input data and negative binomial approach (Love et al., 2014). Further, downstream analysis can be performed to put the differentially expressed genes into a biological context. Gene ontology (GO) term enrichment is another type of commonly used downstream analysis tool that identifies significantly enriched ontologies for three domains: Biological Process, Cellular Components, and Molecular Function. In addition, commercial software Ingenuity Pathway Analysis (IPA) allows identification of canonical pathways and signaling networks that are significantly altered in the analyzed dataset. We carried out GO term analysis of Biological Processes and IPA in Paper III.

3.7 CELL CYCLE ANALYSIS

Cell cycle analysis is a widely used method that utilizes flow cytometry to discriminate cells in different cell cycle phases by measuring DNA content. For univariate analysis, cells are treated with a fluorescent dye, such as propidium iodide (PI) or 4,6-diamidino-2-phenylindole (DAPI), that binds DNA stoichiometrically and the fluorescence intensity thus correlates with the DNA amount they contain. Fluorescence intensities of individual cells are then represented in a histogram allowing identification of relative frequency (percentage) of cells in G_0/G_1, S phase, and G_2/M phases. Multiparameter analysis combining simultaneous measurement of DNA content with other counterstained cellular components or features, such as RNA or proteins, can also be performed by flow cytometry. Another common and especially useful assay is a concurrent measurement of DNA content and 5-bromo-2'-deoxyuridine (BrdU), which is readily incorporated in replicated DNA instead of thymidine. Subsequent immunodetection of BrdU using a specific fluorescent-labeled antibody together with measurement of counterstained DNA allow accurate identification of cells in S phase of the cell cycle. In Paper II, we used bivariate PI stained DNA/BrdU analysis of cell cycle to examine effects of p53 TM on its control of cell-cycle arrest with emphasis on S phase.

Even though the cell cycle analysis by flow cytometry is a very useful method applied in a number of fields, it can have various limitations. PI, which was used in this study, stains RNA in addition to DNA and is unable to pass through plasma membrane. These two disadvantages of PI staining can be easily solved by RNase treatment and permeabilization of the plasma membrane by fixation/detergent treatment. However, in this study, standard fixation step with ethanol was not compatible with GFP reporter expressing cells, as GFP

signal is lost following fixation. To solve this issue, transfected cells were sorted based on the signal of GFP reporter expression and then were immediately subjected to fixation with ethanol.

3.8 GENE AND PROTEIN EXPRESSION

In this thesis, we used reverse transcription quantitative PCR (RT-qPCR) and reverse transcription PCR (RT-PCR) to detect and analyze gene expression in response to different conditions in Paper I and Paper II. In addition, Western blot technique was applied in Paper II to detect proteins of interest in response to different conditions.

3.8.1 Real time qPCR and reverse transcription PCR

Polymerase chain reaction (PCR) is a technique widely used to exponentially amplify a specific segment of DNA or RNA for a broad spectrum of applications. The versatility of PCR has resulted in numerous variants of this technique. RT-qPCR technique is often applied to quantitatively analyze alterations of gene expression levels in real time. The use of fluorescent dyes, such as SYBR® Green, enables monitoring the accumulation of amplified product as the PCR reaction progresses. Then, relative quantification is based on internal reference genes to determine fold-change in expression of the target genes. RT-PCR, another variant of PCR, is commonly used to qualitatively identify gene expression through the synthesis of complementary DNA (cDNA) from purified RNA, which then is amplified by traditional PCR. The simple technique is highly sensitive and produces results rapidly providing sufficient amounts of product that can further be used for sequencing and cloning. The PCR approach has an advantage as it needs only a pair of primers to carry out the amplification and is cost-effective; however, sequence of the target fragment is needed *a priori* to generate primers, which can be a limiting factor. For initial reverse transcription, random hexamers are used to synthesize cDNA from all transcripts present in a sample. In addition, high sensitivity to contamination often leads to the production of misleading results.

Both of these methods were used in paper I in order to quantitatively and qualitatively determine the expression of the different splicing isoforms of *LMNA* and *PLP1* mRNA as a result of TM. RT-qPCR was used in Paper II to determine the impact of p53 TM on gene expression levels of p53 target genes.

3.8.2 Western blot

Western blot or protein immunoblot is an analytical technique extensively used to identify a specific protein of interest within mixture of proteins obtained from e.g. cell lysates and measure relative changes of protein expression between different conditions. In Paper II, two different detection techniques were used to detect and semi-quantify expression levels of various regulators of cell cycle and apoptosis following standard procedures. Direct detection of most proteins was performed using secondary antibodies labeled with IRDye near-infrared (NIR) fluorescent dyes with Odyssey CLx (LI-COR) system. However, we experienced problems detecting p21 enzyme with Odyssey CLx system that could not be resolved, thus

detection of p21 protein in cellular lysates was performed using enhanced chemiluminescence (ECL) technique. Secondary antibodies conjugated with horseradish peroxidase (HRP) that reacts with a detection substrate are used to detect proteins by ECL technique. The light generated during enzymatic reaction is captured on a photographic film, whereas the signal emitted from fluorescent-labeled antibodies is directly captured in the form of light by Odissey CLx (LI-COR) system. Fluorescence detection has gained popularity for its advantages over traditional (ECL) detection. The main drawbacks of ECL system are limited linear and dynamic range of detection resulting in saturated films and underestimation of protein abundance, even though it is more sensitive. In addition, fluorescent western blot detection is time and cost-effective method with a wide dynamic and linear range that offer data reproducibility and quantification.

3.9 MITOCHONDRIAL MEMBRANE POTENTIAL

Mitochondrial membrane potential ($\Delta\Psi m$) is a key indicator of cell health or damage. In healthy cells, $\Delta\Psi m$ is an essential component in the process of ATP synthesis through oxidative phosphorylation. The collapse of the $\Delta\Psi m$ leads to the opening of mitochondrial pores and the subsequent release of cytochrome C into the cytosol triggering downstream events in the apoptotic cascade. Permeant cationic fluorescent dyes is a commonly used tool for measuring $\Delta\Psi m$ to monitor mitochondria function (Perry et al., 2011). Accumulation of fluorescent dyes in the negatively charged mitochondrial matrix of healthy cells is then detected by variety of instruments, thus allowing qualitative or semi-quantitative comparative assessments of the $\Delta\Psi m$ among experimental conditions.

In Paper II, a permeant cationic fluorescent dye, TMRE, was used to monitor changes of the $\Delta\Psi m$ in response to different conditions by flow cytometry. This technique was applied to investigate the impact of p53 TM on the induction of apoptosis. Carbonyl cyanide m-chloro phenyl hydrazine (CCCP), which depolarizes mitochondria by increasing their permeability to protons, was used as a positive control to validate the approach in our experimental system. The main reason for choosing TMRE in this study was its compatibility with our experimental system. Transfected cells express GFP reporter, whereas TMRE emits red fluorescence in healthy cells. In addition, cells with depolarized mitochondria fail to sequester TMRE resulting in diminished levels of red fluorescence, whereas other dyes, such as m-MPI or JC-1, result in fluorescence emission shift, which is incompatible with our experimental setup. In general, other features, such as the lowest mitochondrial binding and inhibition of respiratory electron transport chain, thus being the least toxic to mitochondria, or readily accumulation in active mitochondria, strong fluorescence signal at low concentrations, makes TMRE being preferred for many studies (Nicholls and Ward, 2000).

3.10 CASPASE-3/-7-LIKE ACTIVITY ASSAY

Caspase activity assays with colorimetric or fluorescent output provides simple and convenient means to measure the activity of various activated caspases in cell lysates containing apoptotic cells. In particular, the Caspase-3/-7-like activity assay utilizes a

fluorogenic substrate (N-AcetylAsp-Glu-Val-Asp-7-amino-4-methylcoumarin or Ac-DEVD-AMC). The substrate is cleaved between DEVD and AMC by activated caspases that share the same substrate sequence, such as caspase-3 or caspase-7, releasing highly fluorescent AMC that can be detected using a fluorescence reader (Ex 380 nm / Em 420 - 460 nm) (CST, 2012). Such assays are a quick and convenient method to assess caspase activation in cells. However, cleavage specificities of caspases overlap and identification of a single specific caspase is highly limited as the substrate in a given assay can be recognized by several caspases. Both caspase-3 and -7 are effector caspases responsible for execution of apoptosis. Caspase-3/-7-like activity assay was performed further to investigate effects of p53 TM on the induction of apoptosis in Paper II by measuring activity of effector caspases.

3.11 AGT ACTIVITY ASSAY

AGT activity assay is a well-established sensitive method to detect functional activity of AGT, which repairs O^6-meG, in cell or tissue extracts. Since repair by AGT is a stoichiometric reaction, this method involves measuring [^{3}H]methyl group transfer to AGT protein (Watson and Margison, 2000). Briefly, cellular extracts are incubated with radioactively labeled [^{3}H]-methylated DNA substrate until the reaction is complete. The radioactivity in the protein fraction is measured by liquid scintillation counting and the specific activity can be expressed as fmoles AGT per mg protein in the extracts. In this thesis to show that AGT is indeed suppressed by O^6-benzylguanine (O^6-bzG) at experimental conditions, AGT activity with and without O^6-bzG treatment was assessed in cell extracts. O^6-bzG is a synthetic derivative of guanine that inhibits AGT enzyme and interrupts DNA repair. The main disadvantage of this assay is the use of the radioactive substrate; however, no convenient, highly sensitive alternatives to quantify active AGT have been established.

3.12 CELL VIABILITY ASSAY

Alamar blue assay is a widely used method to study cell viability and cytotoxicity of compounds. When cells are alive, they maintain a reducing intracellular environment that can be directly monitored by a redox indicator Alamar blue. When added to living cells, the active compound resazurin (blue, non-fluorescent) is reduced to resorufin (red, fluorescent) (Lancaster and Fields, 1996). Healthy cells continuously convert resazurin to resorufin, thus increasing overall fluorescence and change in color of the media can be detected by measuring fluorescence (Ex 530-560 nm / Em 590 nm) or absorbance (570 nm and 600 nm), respectively, with the fluorescence detection being more sensitive. This cost-effective assay is similar to traditional tetrazolium reduction (MTT) assay but it is more sensitive, requires fewer steps, offers time-course monitoring and can be combined with other methods, such as caspase activity measurement. The disadvantages of Alamar blue assay are the possibilities of fluorescent interference from compounds being tested and direct toxic effects on the cells (Rampersad, 2012). The Alamar blue assay was used to in Paper II to evaluate cytotoxicity of inhibitor O^6-bzG.

4 RESULTS

This section provides a summary of the main findings in Papers I-III. For details and comprehensive discussions, the reader is referred to the respective publication.

4.1 PAPER I: TRANSCRIPTIONAL MUTAGENESIS REDUCES SPLICING FIDELITY IN MAMMALIAN CELLS

Knowledge about effects and consequences of transcriptional errors on splicing fidelity is scarce in general. Analysis of RNA sequencing data on RNA Pol II fidelity has suggested that transcriptional errors have the potential to significantly affect splicing fidelity and may contribute to disease development (Carey, 2015). Notably, the impact of TM on splicing fidelity had not yet been investigated. We hypothesized that TM events induced by a DNA lesion could change the recognition specificity of a splicing signal in pre-mRNA activating an alternative splice site and subsequently resulting in aberrant splicing in mammalian cells. With this study, we aimed to investigate the effects of both O^6-meG and 8-oxoG induced TM on splicing fidelity by assessing activation of alternative splicing using *LMNA* and *PLP1* minigene constructs as probes (Figure 7).

Firstly, we constructed two non-replicative minigene splicing reporter systems using the probe genes, *LMNA* and *PLP1*, with established links to severe human diseases caused by aberrant splicing due to a single base substitution within each gene. Each reporter system consisted of three plasmids. One of the control plasmids contained either wild type *LMNA* or *PLP1* sequences and the other contained the mutated sequences with the specific c.1284C>T or c.347C>A substitutions. The third plasmid contained either a site-specific O^6-meG at codon 608 for *LMNA* gene or 8-oxoG at codon 116 for *PLP1*. The presence of O^6-meG and 8-oxoG was confirmed either by sequencing of the plasmids or by an Fpg nicking assay.

Next, mammalian cells (HEK293, WI-38, wild type and *OGG1*$^{-/-}$ MEFs) were transiently transfected or nucleofected with the reporter constructs to determine the extent of TM and the effects of TM on splicing fidelity. In order to determine levels of TM induced by both lesions, cellular transcripts from the damaged reporter were analyzed by sequencing of RT-PCR products. Our results demonstrated that the presence of O^6-meG and 8-oxoG on the transcribed strand induced TM in these cells to similar extent ranging 1–4%. Moreover, levels of TM increased by one order of magnitude when repair was compromised either by inhibition of AGT or *OGG1* deficiency (20% in HEK293 induced by O^6-meG and 40% in *OGG1*$^{-/-}$ MEFs by 8-oxoG).

Further, we investigated to what extent these misincorporation events in pre-mRNA affected splicing fidelity since they changed the recognition specificity of the splicing signal. As described above, these minigene reporters were constructed in a way that GFP was only expressed if alternative splicing occurred. Thus, fluorescence intensities were measured with flow cytometer and revealed that lesion-induced TM indeed activated alternative 5′ splice sites for both probe proteins (Figure 9). Notably, TM activated alternative splicing in HEK293 with active DNA repair, but not in WI-38 or MEFs. As expected, the levels of

alternative splicing increased substantially in all cell types when DNA repair of each lesion was impaired, which is in agreement with the observed higher frequency of TM.

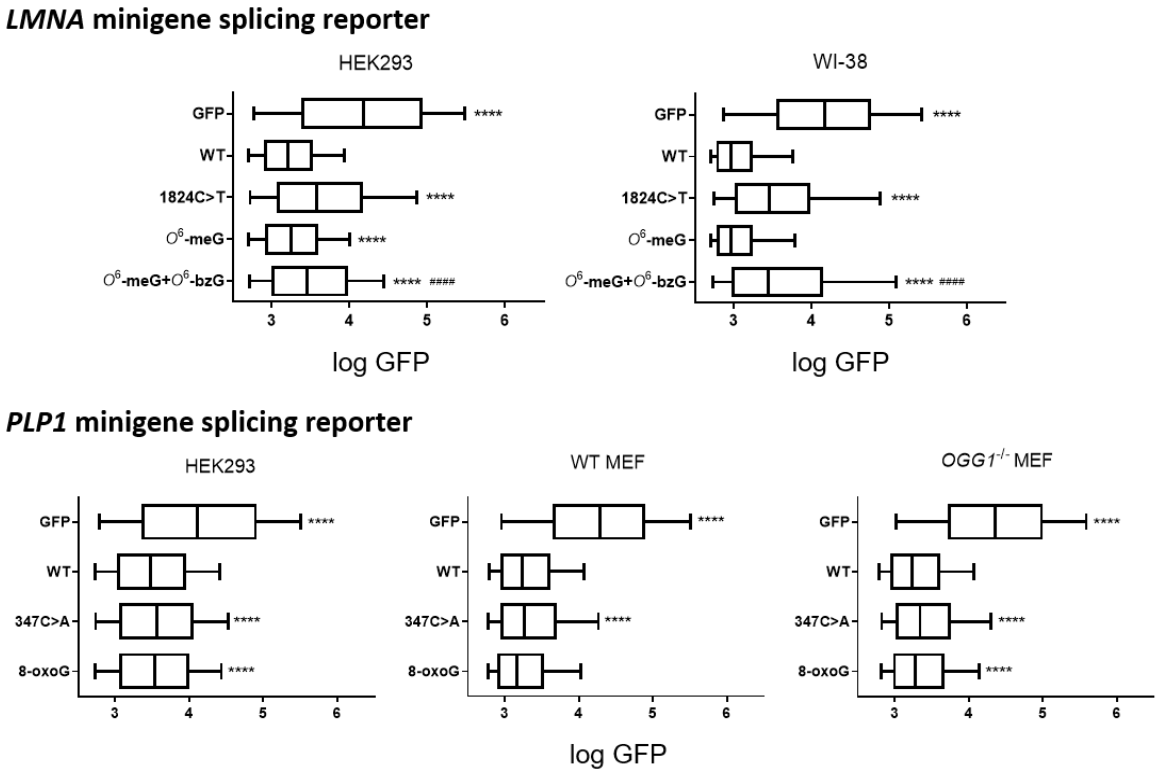

*Figure 9. Box plots of distributions of fluorescence intensity in different cells transfected with the LMNA or PLP1 minigene reporters. **** in comparison to the WT reporter, #### in comparison to the O^6-meG reporter. Image adapted from Paredes et al., 2017.*

In addition, flow cytometry results were further confirmed by quantitative and qualitative PCR. With this approach wild type LMNA and progerin transcripts were detected in both HEK293 and WI-38 cells with active or inactive DNA repair, the latter being substantially higher, as expected. This further confirmed that O^6-meG-induced TM affected splicing of *LMNA* pre-mRNA by activation of the cryptic splice site. In case of *PLP1* reporter, where both PLP1 and DM20 splicing isoforms can be expressed simultaneously under physiological conditions, the effects of 8-oxoG-induced TM on splicing fidelity were not as clear. Such effects were reflected by changes in the ratio between the two isoforms. Semi-quantitative PCR analysis was carried out and revealed that 8-oxoG-induced TM affected splicing fidelity of *PLP1* pre-mRNA by increasing the selection of alternative splicing site in all cell types, except for wild type MEFs.

To summarize, the results of this study showed that the presence of O^6-meG and 8-oxoG on the template strand of a gene induces TM if left unrepaired. This is in agreement with previous *in vivo* studies and extends the knowledge about the impact of DNA damage on transcription fidelity. We also demonstrated for the first time that lesion-induced transcriptional misincorporations could change the recognition and selection patterns of splicing sites, thus activating alternative splicing and resulting in the production of alternatively spliced RNA isoforms in mammalian cells. Our results indicate, TM may not

only change translational outcome of a gene by affecting the coding specificity of a codon, but also by changing regulatory sequences, such as splicing signals, and thereby altering the recognition of splice sites. The present study indicates that TM could contribute significantly to the burden of mutant proteins within a cell by reducing splicing fidelity, thereby inducing adverse effects on cellular homeostasis.

4.2 PAPER II: *O*^{6}-METHYLGUANINE-INDUCED TRANCRIPTIONAL MUTAGENESIS REDUCES P53 TUMOR-SUPPRESSOR FUNCTION

Previous studies have reported that mutated transcripts may be produced due to lesion-induced TM *in vivo* and that these transcripts can be translated into proteins with altered functions (Bregeon et al., 2009; Burns et al., 2010; Saxowsky et al., 2008). At the time this study was conducted, there was only one previous study reporting biologically relevant effects of TM in mammalian cells. A study in murine cells demonstrated that 8-oxoG bypass during transcription resulted in activation of RAS protein and subsequent MAPK downstream signaling (Saxowsky et al., 2008). This demonstrated the ability of TM to activate oncogenic signaling pathway and suggesting a link between TM and tumorigenesis. With this in mind, we aimed to investigate the biologically relevant effects of TM induced by *O*^{6}-meG on protein function and downstream signaling using p53 tumor suppressor as a probe protein and with the focus on two main functions of p53: cell cycle regulation and apoptosis.

Firstly, a non-replicative TM reporter plasmid containing *TP53* gene with a site-specific *O*^{6}-meG placed in codon 248 and *GFP* gene for an internal transfection control was constructed. Two plasmids encoding either wild type p53 or mutant p53 with c.742C>T (R248W) substitution were additionally constructed to serve as controls for cellular responses to the expression of p53 variants. The presence of *O*^{6}-meG was confirmed employing restriction enzyme assay based on the fact that presence of *O*^{6}-meG blocks the cleavage.

To determine the levels of p53 TM and to study the effects of p53 TM on cellular homeostasis, we transiently transfected human p53-null H1299 cells either treated or untreated with *O*^{6}-bzG, thereby also allowing us to examine the impact of DNA repair. This particular cell line was chosen in this study to avoid any interference from endogenous levels of p53 allowing the interpretation of obtained data. Moreover, H1299 cells are common model to study the effects of p53 mutagenesis (Willis et al., 2004). RNA sequencing revealed that *O*^{6}-meG induced very low levels of uridine misincorporation (0.14%) at codon 248 in p53 transcripts from H1299 cells, but increased hundred-fold (14.7%) when DNA repair was impaired.

We continued to investigate if expression of mutated p53 transcripts due to TM induced by *O*^{6}-meG would produce p53 proteins with altered functions. qPCR analysis of several p53 target genes, including *CDKN1A* and *BBC3*, was performed and showed that the transactivation capacity of p53 was significantly reduced due to TM. The obtained gene expression results strongly suggested that TM at codon 248 deregulated p53-mediated cell cycle control by disrupting the p21-regulated functional G1/S checkpoint. In addition, gene

expression data also indicated that TM of p53 reduced the capability to activate the intrinsic apoptosis pathway.

Results from immunoblotting confirmed that O^6-meG induced TM of p53 reduced protein levels of p21. Previous studies have shown that p21 is crucial for p53-dependent control of cell cycle arrest and for the effective tumor suppressor function of p53 (el-Deiry et al., 1994). Loss of functional G1/S checkpoint due to p53 TM was further verified by deregulated levels of other G1/S checkpoint regulatory proteins. In addition, disrupted control of G1/S checkpoint was confirmed by bivariate cell cycle analysis. Results showed that O^6-meG-induced TM of p53 reduced the tumor suppressor function to arrest the cell cycle by allowing a large population of cells to transit into S phase.

Ultimately, we checked the effects of p53 TM induced by O^6-meG on the potential to induce apoptosis. The consequences of p53 TM on two hallmarks of apoptosis, loss of mitochondrial membrane potential and caspase activation, were evaluated. Results revealed that TM impaired the induction of apoptosis as was shown by a reduced loss of mitochondrial membrane potential and reduced level of caspase activation, thus attenuating the effective function of p53 as a tumor suppressor.

To summarize, our results showed that the presence of O^6-meG on the template strand of *TP53* gene can induce TM in human cells when AGT is lacking. More importantly, the level of expression of mutant p53 protein from the resulting mutated transcripts was sufficient to attenuate the tumor suppressor function and disrupt downstream signaling that is crucial for functional cell cycle checkpoints and apoptosis. As summarized in Figure 10, findings of the present study shed light on the consequences of lesion-induced TM and provide evidence that TM could have detrimental consequences on cellular homeostasis supporting the proposed contributions of TM to human disease development, and especially tumorigenesis.

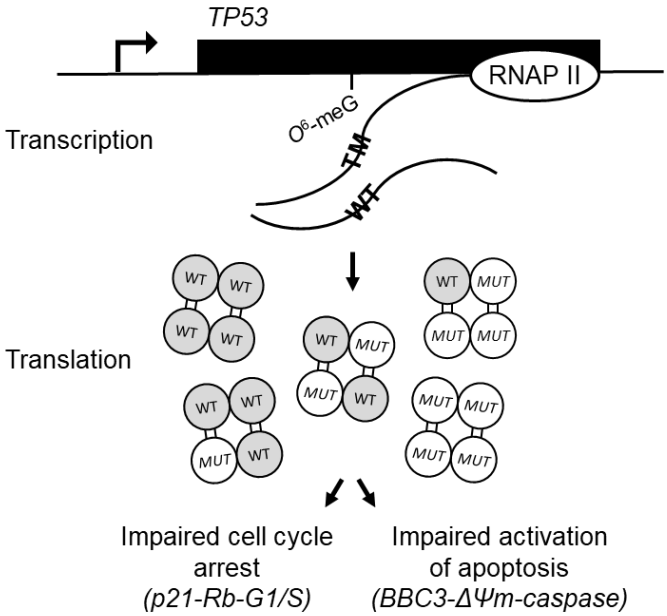

Figure 10. O^6-meG-induced transcriptional mutagenesis can reduce p53 tumor suppressor function by impairing control of cell cycle arrest and apoptosis. Image reused from Ezerskyte et al., 2018.

4.3 PAPER III: TRANCRIPTIONAL MUTAGENESIS DRAMATICALLY ALTERS GENOME-WIDE P53 TRANSACTIVATION LANDSCAPE

In Paper II, we have demonstrated that O^6-meG-induced TM of p53 reduced its tumor suppressor function, which is exerted mainly through transcriptional activation of numerous target genes. We examined the effects of TM on p53 transactivation capacity only on a limited number of target genes, but p53 has been shown to regulate expression of a few hundred genes related to numerous cellular processes to maintain homeostasis. In this study, we aimed to determine time-dependent formation of TM induced by O^6-meG in p53 transcripts and further to expand the assessment of the effects of TM on p53 transactivation of target genes at a genome-wide level.

We used the same expression constructs containing a *TP53* gene with a site-specific O^6-meG and p53-null H1299 cells as in Paper II. Human H1299 cells either treated or untreated with O^6-bzG were transiently transfected and RNA sequencing was carried out on samples from an early (6 h) and late (24 h) time points. First, we determined the extent of O^6-meG-induced TM in p53 transcripts at codon 248 at the two time points. In agreement with previous study, results of RNA-seq revealed that transcription past O^6-meG in cells with active AGT resulted in low levels of uridine incorporation opposite the lesion. In contrast, levels of uridine misincorporation increased by two orders of magnitude when AGT was inhibited at both time points. Furthermore, levels of TM at the early time point were almost 3-fold higher than at the later time point (47% of p53 transcripts contained uridine at 6 h, whereas at 24 h only 18%). High levels of uridine misincorporation at the early time point were further confirmed using a PCR-based approach.

Next, differential gene expression analysis was further performed to investigate if there were any p53 status-dependent differentially expressed genes (DEGs). Results revealed that the total number of DEGs increased over-time in response to all vectors when compared to the control without expression of p53, except for the R248W mutant p53, which did not induce any DEGs (not including *TP53* expressed from the plasmid).

Furthermore, we measured transactivation levels of 346 high-confidence p53 target genes (HCGs) previously described in 16 genome-wide data sets and one literature-based data set that met the criterion of a target gene (Fischer, 2017). TM of p53 reduced expression levels of 97.6% and 64% of the wild type p53 induced HCGs at early and late time points, respectively. Moreover, time-dependent effects of TM on p53-dependent transactivation of HCGs were observed revealing three clusters: early repression (*e.g. BAX* was repressed only at the early time point), continuous repression (*e.g. CDKN1A* was repressed at both time points), and late repression (*e.g. CAV1* was repressed only at the late time point).

In order to further elucidate which cellular processes were affected by p53 TM in time-dependent manner, we sought to perform gene ontology (GO) term analysis of Biological Processes associated with the HCGs identified in each cluster. As expected, genes annotated as being involved in p53 signaling and apoptotic signaling were represented in all three clusters. Furthermore, cellular processes regulated by the genes identified in both early and

continuous repression clusters clearly overlapped and included GO terms related to DNA damage response (*e.g.* Signal transduction in response to DNA damage) and regulation of cell cycle (*e.g.* Cell cycle checkpoints) (Figure 11). In agreement with the TM reduced transactivation of *CDKN1A* at both time points, repression of transcriptional downregulation of numerous genes was observed at the late time point, but no GO term enrichments could be found. Pathway analysis using IPA revealed that the E2F transcription factor family was among the most important upstream regulators of these genes.

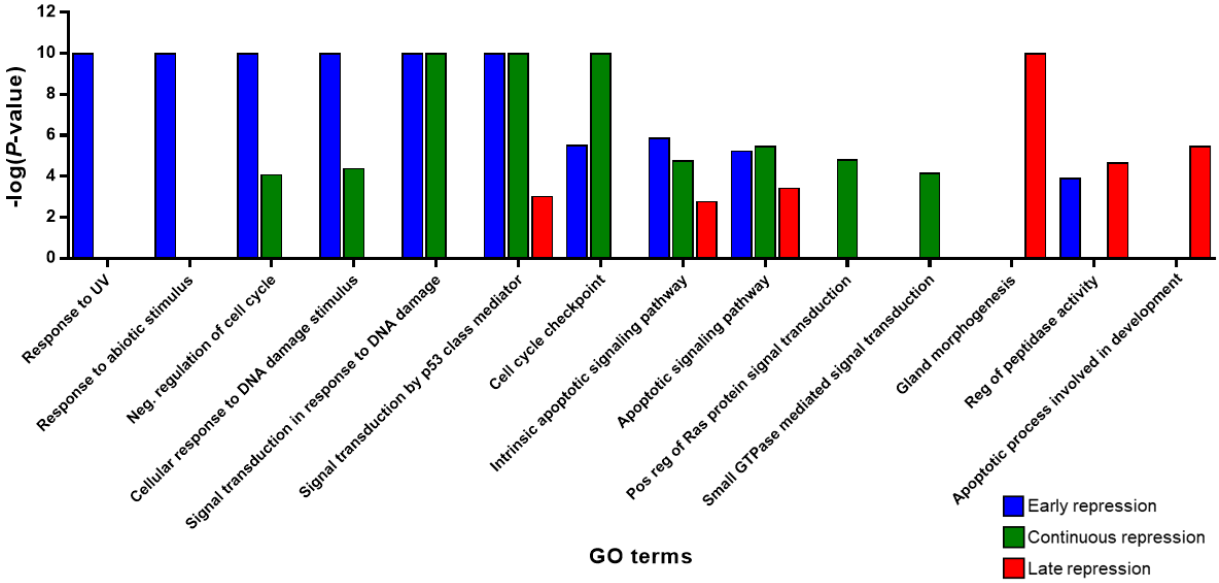

Figure 11. Over-represented GO terms in each temporal cluster. Image adapted from Ezerskyte et al., manuscript.

To summarize, we showed that TM of p53 deregulated the transactivation or downregulation of numerous target genes involved in several canonical signaling pathways in a time-dependent manner. All these pathways are crucial for the effective p53 tumor suppressor function and maintenance of cellular homeostasis. Overall, these findings further support the possible contribution of TM in the multistep process of tumor development.

5 GENERAL DISCUSSION

Undamaged DNA template is required for both essential cellular processes – DNA replication and transcription. Despite numerous repair mechanisms, both enzymes can encounter DNA lesions on DNA templates resulting in mutations. Mechanisms and consequences of mutagenesis from DNA replication past DNA damage are well characterized and have been shown to be highly important for genetic variability, evolution, hereditary diseases and tumorigenesis in mammalian systems (Bregeon et al., 2009; Lynch, 2010). DNA lesion-induced TM has the potential to produce a population of mutated transcripts as long as the lesion remains unrepaired, which subsequently can be translated into mutant proteins. Thus, TM can adversely affect cellular homeostasis. The extent to which lesion-induced TM influences protein function and the role of these seemingly transient changes in cellular homeostasis in disease development remains unclear (Figure 12) and was one of the questions this thesis wanted to address.

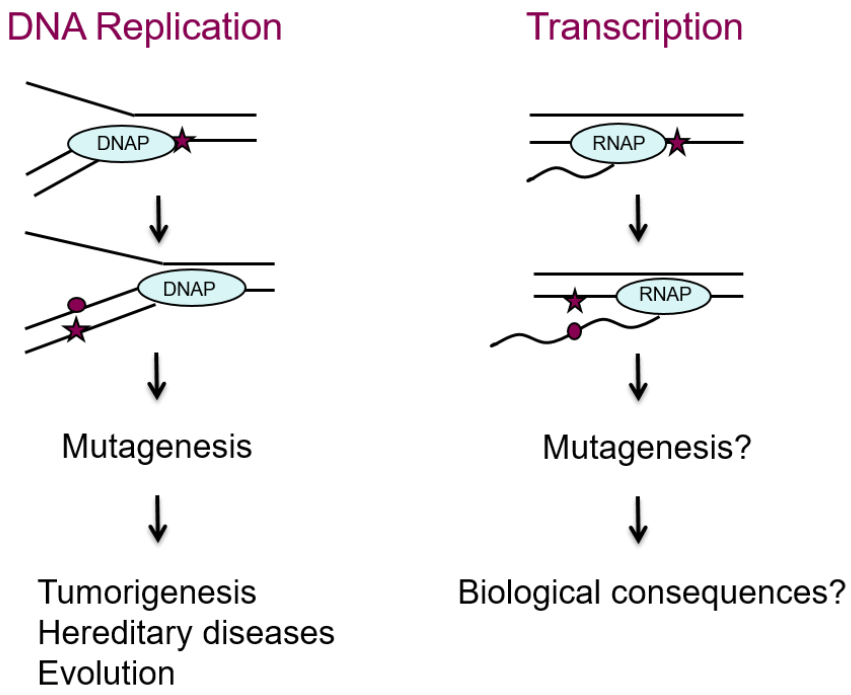

Figure 12. Schematic presentation of the consequences of mutagenic bypass of DNA lesion by elongating DNA and RNA Pols.

Issues with methodologies

Gradual loss of proper homeostasis is a reason of many diseases as well as aging (Hartl, 2016). A pool of mutant proteins can be produced due to lesion-induced TM and thereby substantially contribute to the burden of aberrant proteins in a cell with potentially adverse effects on cellular homeostasis. The impact of mutagenesis during transcription (TM) may not have received an adequate attention due to several reasons. TM induced errors are presumably transient and present in only a fraction of the proteins produced. They are moreover inheritable, so detection and quantification of these errors are technically challenging and require specialized methods, thus limiting studies of TM.

Several approaches have been developed to study the effects of DNA lesions on elongating RNA Pol during transcription and to investigate the potentially detrimental effects on cellular homeostasis. The first *in vitro* studies used RNA pols purified from various sources with DNA templates containing site-specific DNA lesions and revealed that TM indeed occurs. Most of the early studies focused on the effects of bulky helix-distorting DNA lesions that generally arrest elongating RNA Pols which led to the discovery of TC-NER (Mellon et al., 1987). Later, to study effects of TM on protein function *in vivo*, protocols were developed to produce non-replicative reporter constructs that are introduced into bacterial or mammalian cells and contained a site-specific DNA lesion in a gene of a protein with biochemically detectable activity (Bregeon et al., 2003; Bregeon and Doetsch, 2004; Shaw et al., 2002; You et al., 2000). However, this approach does not provide information on possible effects on cellular homeostasis. To investigate how lesion-induced TM could affect cellular homeostasis, DNA lesions were instead placed at defined positions in gene sequences that would result in proteins harboring mutations associated with disease phenotypes and which function could be biochemically measured (Ezerskyte et al., 2018; Saxowsky et al., 2008). Reporter plasmids containing site-specific DNA lesions in target genes is a powerful and informative tool allowing to examine effects of TM induced by any DNA lesion in any sequence of interest. However, the most limiting step in this system is the production of large amounts of reporter constructs containing DNA lesions at defined positions as plasmids have to be generated *in vitro*. Although it is theoretically feasible to introduce any DNA lesion into transcription templates, such protocols are time-consuming and technically challenging, thus resulting in limited yields of pure plasmids suitable for transfections.

Although experiments investigating individual loci or using reporter constructs have provided a foundation for the study of TM, they do not represent a genome-wide extent of TM. To establish the role of TM in disease development, especially tumorigenesis, it is important to determine the genome-wide extent of these events in mammalian cells under various conditions. Now, next-generation RNA sequencing technology enables analysis of the entire cellular transcriptome of any organism and provides another approach to determine multiple facets of TM. However, standard protocols cannot accurately determine true levels of transcriptional errors on genome-wide level because of the high error rates of cDNA synthesis and the sequencing reaction. As a result, protocols were improved to overcome these limitations and to enable sensitive determination of transcriptional errors induced by both the inherent RNA Pol II infidelity and by DNA lesions throughout the transcriptome of any organism. Several novel approaches, including the mining of RNA-seq data for splicing errors (Carey, 2015), the high-resolution sequencing method (Imashimizu et al., 2013), the replicated sequencing method (Gout et al., 2013), and the circle-sequencing method (Acevedo and Andino, 2014), with different advantages and drawbacks have revolutionized the field allowing determination of the extent of both spontaneous transcriptional errors due to RNA Pol II infidelity and misincorporations due to TM at the genome-wide level (for an review of these methods, see (Gordon et al., 2015)). Indeed, a recent study presented a novel method, referred to as ARC-seq, which was developed to detect misincorporation events due

to lesion-induced TM with high accuracy, and applied it to determine if oxidative stress elevated levels of TM in yeast (Reid-Bayliss and Loeb, 2017).

Effects of TM

After confirmation that also TM can result in the production of proteins with altered functions *in vivo*, a possible role of TM in aging and pathogenesis of various human diseases, including cancer and neurodegeneration, has been proposed (Bregeon et al., 2003; Saxowsky and Doetsch, 2006). Increasing lines of evidence suggest that TM could be a mechanism which contributes to various stages of tumorigenesis by inactivation of tumor suppressors or activation of oncogenes that could potentially stimulate proliferation of an initiated neoplastic cell or allow it to escape growth-inhibitory signals (Brégeon and Doetsch, 2011; Morreall et al., 2013; Saxowsky and Doetsch, 2006). Somatic mutations, which are detected in human tumors, are consistent with those occurring due to lesion-induced TM, supporting the hypothesis that TM is one possible route of tumorigenesis. These mutations can lead to the constitutive activation of proliferation and survival signaling by inactivating negative regulators of proliferation, including p53, or activating oncogenes, such as RAS (Morreall et al., 2013). As shown in Figure 13, mutations in *e.g. RAS* and *TP53* genes are key genetic alterations driving progression of colorectal cancer (Markowitz and Bertagnolli, 2009). Although transient, the deregulation of these key regulators due to TM could provide the initiated neoplastic cell with growth and survival advantage and possibly result in the fixation of the mutation in the genome following replication if the DNA lesion remains unrepaired. Indeed, this concept of retromutagenesis has previously been proposed but is yet to be shown in mammalian cells (Brégeon and Doetsch, 2011; Morreall et al., 2015).

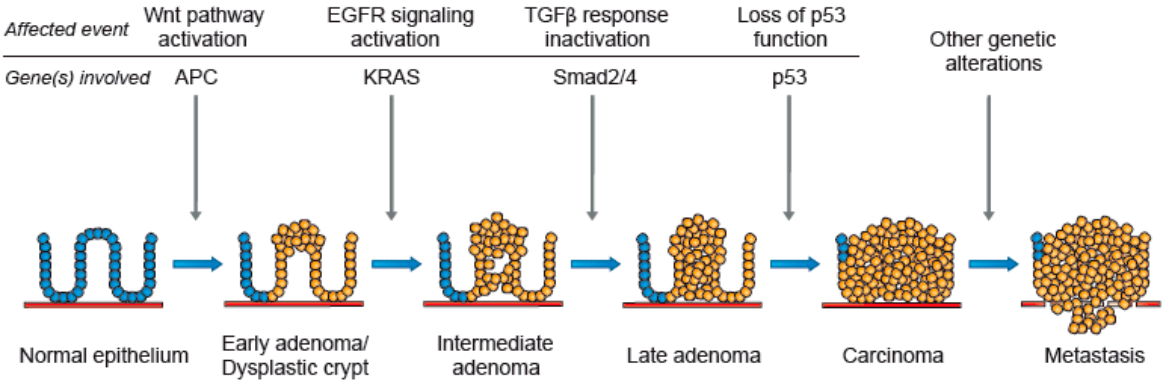

Figure 13. Classical pathway of colorectal cancer progression from normal epithelium. Image adapted from Davies et al. 2005, with permission from Springer Nature.

In addition, it has previously been suggested that TM could be a contributing mechanism to anomalies during development, especially neurodevelopmental deficits, where exposures to DNA damaging agents could have devastating effects due to TM on transcriptional outcome of genes required for normal development and function (Saxowsky and Doetsch, 2006; Wells et al., 2009). One could also speculate that DNA lesion-induced TM could affect spatial and temporal gene expression patterns leading to disturbances of precisely orchestrated p53-dependent apoptosis during development with detrimental consequences. This is, however,

still to be shown. Besides, TM is also implicated to play a role in the etiology of neurodegenerative disorders that are characterized by aggregates of misfolded proteins, such as Parkinson's and Alzheimer's diseases (Basu et al., 2015). Analysis of patient samples revealed that altered proteins originate from mutated RNA transcripts without detectable mutations in genomic DNA indicating that mutations within proteins have arisen as errors during transcription due to increased oxidative DNA damage in ageing brain (van Leeuwen et al., 1998).

Notably, despite the proposed contributions in pathogenesis of various diseases, the potential of transient TM errors due to unrepaired lesions to affect cellular homeostasis is virtually unknown. To date only a few studies have investigated the impact of TM on cellular homeostasis and provided direct evidence that TM can lead to synthesis of proteins with changed function that are able to activate oncogenic signaling pathway or affect splicing fidelity by changing the recognition of the splice site (Ezerskyte et al., 2018; Paredes et al., 2017; Saxowsky et al., 2008). The first direct evidence supporting the involvement of TM in tumorigenesis came from a study using a non-replicating construct containing a site-specific 8-oxoG lesion in the transcribed strand of the *RAS* gene. Bypass of the lesion resulted in approximately 14% of the transcripts being mutated in *OGG1*$^{-/-}$ MEFs, which resulted in production of constitutively active mutant (Q61K) RAS protein. This was sufficient to activate downstream effector ERK, thus demonstrating the ability of TM to initiate oncogenic signaling pathway (Saxowsky et al., 2008). Our study in Papers II-III further confirmed the role of TM in tumorigenesis. We showed that O^6-meG-induced TM resulted in 14.7% of mutated transcripts in human cells with impaired DNA repair and that the subsequent expression level of mutant R248W p53 was sufficient to impair tumor suppressor function of p53 by deregulating cell cycle control and activation of apoptosis. Further investigation at a genome-wide level revealed that TM of p53 deregulated expression of numerous p53 target genes.

As discussed above, alternative splicing is a highly regulated process that requires extreme precision. Effects of splicing infidelity can have detrimental biological consequences that are reflected by a large number of severe diseases associated with aberrant splicing (Ward and Cooper, 2010). The ability of spontaneous RNA Pol errors to significantly affect splicing fidelity has been proposed before (Doetsch, 2002; Fox-Walsh and Hertel, 2009) and recently suggested from RNA-seq data mining analysis (Carey, 2015), but never tested *in vivo*. In addition, the impact of TM on splicing fidelity has not yet been investigated. We have revealed for the first time that DNA lesion-induced TM can lead to specific mutations in pre-mRNA molecules that deregulate splicing thus reducing splicing fidelity in mammalian cells. This indicates that TM could contribute to the mutant protein burden within the cells affecting cellular homeostasis; however, the extent and the effects of these events on cellular homeostasis are unknown and require further investigation. In certain pathogenesis, such as tumor progression, alterations in alternative splicing patterns can be a critical component for disease development and have been found to drive cancer (Carstens et al., 1997; Sveen et al., 2016). This emphasizes the impact of lesion-induced TM on splicing fidelity resulting in

aberrant splicing and supports TM as a contributing mechanism in tumorigenesis. In addition, accurate transcription is necessary for normal development (Wells et al., 2009). Reduced splicing fidelity in response to DNA lesions due to TM could possibly have detrimental consequences at any developmental stage.

Interesting observations

The results of the present work are summarized and discussed in detail in their respective articles. Nonetheless, two interesting observations in the present work are worth discussing further. First, several studies have shown that dominant-negative p53 mutants, such as the R248W p53, are capable of altering gene expression patterns not simply due to loss of function of wild type p53, but also due to gain of new functions. Thus, in addition to loss of function and dominant-negative effects, it has been proposed that several p53 mutants can exhibit oncogenic gain of function (GOF) phenotypes that promote tumorigenesis (O'Farrell et al., 2004; Scian et al., 2004; Strano et al., 2007; Willis et al., 2004). Several mechanisms for GOF phenotype, which is characterized by enhanced proliferation, resistance to anticancer treatment, increased invasiveness, migration, and genomic instability, have been proposed (Strano et al., 2007). Some studies have proposed that mutant p53 exhibit changed binding preferences for target response elements compared to wild-type p53, through which it regulates numerous genes, allowing transactivation of different genes (Resnick and Inga, 2003). While others propose that mutant p53 can interact, sequester, and inactivate other proteins and change their transactivation ability resulting in altered gene expression patterns (Strano et al., 2007). However, in addition to scarce knowledge of the GOF molecular mechanism, evidence for additional oncogenic properties of different p53 mutants remains inconsistent. Recent genome-wide approaches showed no evidence for gain of function phenotype of dominant-negative p53 mutants in patient-derived lymphocytes (Zerdoumi et al., 2017). This is further supported by our RNA-seq data presented in Paper III, where we could not detect any mutant-specific differentially expressed genes in response to the R248 p53 expression in H1299 cells compared to both the control (no p53 expression) and the wild type p53 at both early and late time points. Contradictory results may have arisen due to nature of earlier studies that greatly relied on microarray hybridization analysis and chromatin immunoprecipitation (ChIP) assays. Thus, high-throughput RNA-seq analysis of gene expression profiles of different p53 mutants is required and will provide more insight about mutant p53 capabilities to specifically alter gene expression patterns.

Another interesting observation is the time-dependent decrease of mutated transcripts with uridine misincorporation induced by O^6-meG in H1299 cells when AGT repair is compromised (Paper III). We reported that 47% of p53 transcripts contained uridine at 6 h, whereas only 18% were mutated at 24 h, the latter being in agreement with our previous study (Ezerskyte et al., 2018). To assure that the surprisingly high level of TM at the early time point is not an artifact or a technical experiment mistake, we further confirmed this by another PCR-based approach, which showed similar 51% uridine incorporation opposite the lesion in cells with inactive AGT. In addition, we confirmed the inhibition potential of O^6-bzG and showed that more than 95% of the AGT activity was inhibited in the presence of O^6-

bzG (10 µM) for up to 24 h. This eliminated the possible explanation of insufficient inhibition resulting in repair of the lesion by residues of active AGT. Furthermore, we also considered if the elimination of a damaged plasmid due to cell death could explain this finding. Increased levels of apoptosis were observed in cells expressing wild-type p53. In this case, an increase of TM levels could actually be expected, as the wild type p53 expressing cells from the damaged plasmid would induce apoptosis and die, leaving the cells in the culture that express mutant p53 due to TM. After considering all, this might indicate that other repair pathways, such as BER or NER could play a backup role for AGT or even yet an unidentified homologue of AGT in human cells could repair the lesion. Indeed, an *in vitro* study have shown that both *E. coli* and human excinucleases belonging to NER pathway were able to remove O^6-meG, however with low efficiencies (Huang et al., 1994), although this pathway is primarily thought to function in the repair of DNA-helix distorting lesions. Moreover, several alkylation adducts are recognized and repaired by the BER pathway (Kaina et al., 2007). Interestingly, *E. coli* possesses two alkyltransferases Ada and Ogt with slightly different preferences for substrate; however, both are capable to eliminate O^6-meG lesion indicating that human cells could also possibly have homologue proteins able to repair O^6-meG in addition to AGT, but yet to be discovered (Sassanfar et al., 1991). This is supported by the fact that residual activity of AGT was detected in *MGMT*$^{-/-}$ mice liver and MEFs, however these mice displayed no obvious phenotypic or pathological abnormalities suggesting the existence of additional methyltransferase activity (Glassner et al., 1999). In addition, alkyltransferase-like proteins (ATLs), so far identified in prokaryotes and lower eukaryotes, have been recently revealed by *in silico* analysis with functional motif similarities to AGT and abilities to bind alkylated DNA and to trigger the NER pathway (Margison et al., 2007; Tubbs et al., 2009). Additionally, it has been suggested that TC-NER could possibly play an auxiliary role in the clearance of this lesion as O^6-meG partially blocks RNA Pol II, however, this needs to be further investigated (Dimitri et al., 2008).

To summarize, the work presented in this thesis sheds some light on the impact of TM on cellular homeostasis, which may have been initially underestimated due to transient nature of the transcripts and challenging approaches of detection and quantification. However, many issues have yet to be addressed and further thorough investigation is required to fill the gap in scientific knowledge about possible mechanisms by which TM might contribute to human disease development. This can be achieved by combined use of both reporter-based DNA templates containing site-specific DNA lesions and RNA-seq.

6 CONCLUSIONS AND FUTURE OUTLOOK

The extent to which lesion-induced transcriptional mutagenesis (TM) influences cellular homeostasis and its contributions to development of human diseases is not well understood. A number of studies have shown that various DNA lesions can induce misincorporations during transcription and generate mutated transcripts. With the present work, we wanted to shed some light on the biological consequences of TM and provide evidence that lesion-induced TM can adversely affect cellular homeostasis and require further investigation.

In agreement with previous studies, we conclude that both 8-oxoG and O^6-meG present on the template strand of an actively transcribed gene are mutagenic during transcription if left unrepaired and instruct the misincorporation of A and U opposite the lesion to a high degree *in vivo* (Paper I-III).

We show for the first time that lesion-induced TM at a splice site can significantly reduce splicing fidelity *in vivo* leading to the production of disease associated splice forms and/or disrupting ratios between alternatively spliced isoforms. (Paper I).

We show that expression of the dominant-negative R248W mutant p53 due to O^6-meG-induced TM in human cells is sufficient to significantly reduce p53 tumor suppressor function, resulting in deregulated downstream signaling. We show that attenuation of p53 function due to TM impairs its control of cell cycle checkpoints and induction of apoptosis (Paper II). So far, it is the second publication showing biologically relevant effects of TM on cellular homeostasis strongly suggesting TM as a contributing mechanism in tumorigenesis by inactivation of tumor suppressors or activating oncogenes.

We conclude that O^6-meG-induced TM of p53, on a genome-wide level deregulates the transactivation or downregulation of many target genes involved in the regulation of cellular processes, such as cell-cycle arrest, apoptosis, and DNA damage response, all of which are crucial for its tumor suppressor function (Paper III).

Even though TM has been proposed to be involved in pathogenesis of various human diseases, only a few studies recently started to shed light on biologically relevant consequences of lesion-induced TM *in vivo*. The work presented in this thesis together with the results from other studies strongly suggest a possible role of TM in the multistep process of tumorigenesis by impairing the function of proteins, such as tumor suppressors or oncogenes, which are crucial for controlling cellular homeostasis in an already initiated neoplastic cell. In general, it seems that the effects of these transient and inheritable RNA mutations on cellular homeostasis may have been underestimated. Furthermore, in recent years, next-generation sequencing approaches have allowed studies of transcription fidelity on a genome-wide level, confirming adverse effects on protein function and cellular homeostasis. All these experiments together reveal a previously unappreciated role of transcriptional errors due to both RNA Pol infidelity and TM in deregulating cellular homeostasis and present new mechanisms by which cells can acquire disease phenotypes. To date there are no genome-wide studies on transcriptional errors in mammalian cells, thus the

extent of transcription errors as wells as the consequences of these events to the cellular homeostasis in higher organisms are unknown.

More research within this field is highly necessary to establish mechanisms of how TM could contribute to disease development. The genome-wide RNA-seq technology offers new opportunities to study this virtually untouched field in-depth, which will provide researchers with invaluable data and possibly will lead to intriguing discoveries. Of high interest, RNA-seq could be used to determine extent of TM in response to various genotoxic agents and in cells or tissues with different DNA repair capacities. Worth repeating that terminally differentiated cells have attenuated repair capacities, but rely on transcription to maintain their homeostasis. Moreover, intriguing results could be expected from various tumor samples of patients or model organisms and monitoring of the extent of TM during different tumor development stages. However, a major challenge of the genome-wide RNA-seq will be to connect these errors directly to the changes in cellular function and monitor their effect on cellular health. Even though RNA-seq provides a powerful tool to study the genome-wide extent of TM in any organism under any conditions, the more traditional transcription template approach, as was used here, remains the extremely useful and informative tool. Primarily because this approach allows to investigate effects of a specific DNA lesion on transcription machinery and to evaluate these effects specifically on cellular homeostasis *in vivo*. The impact on elongating RNA Pol II has been investigated only for a small number of lesions. Given the variety of lesions that can be induced in an organism by a plethora of agents, extensive research is further required to shed light on possible perturbations to RNA synthesis exerted by other DNA lesions.

7 ACKNOWLEDGEMENTS

I would like to thank my principal supervisor, Docent **Kristian Dreij**, for giving me this opportunity to join his group as a PhD student. Thank you so much for all your support and guidance during my studies, constant supervision and always being available, I have learned so much from you. Most importantly, after meeting so many students, who have experienced variety of problems with their supervisors, I can conclude that I was very lucky to have you as my supervisor. Thank you!

I really would like to thank Docent **Magnus Olsson** for being my co-supervisor. I really appreciate your input for my second publication. You might think that it was just a small contribution, but your expertize within the field of apoptosis and suggestions allowed me to finalize the second manuscript after being stuck for so long that I even thought I might not finish my studies at all.

I would also like to acknowledge Prof **Ulla Stenius** for being my co-supervisor and an inspiration for women in science.

A special thanks to my dear mentor, Prof **Ulf Rannug**, who was constantly supporting and motivating me during this period of time. Thank you for your advices and for being always available when I needed any help, inspiration or just to share my worries or thoughts. It means so much for me!

I would like acknowledge our collaborators **Prof David A. Scicchitano, John A. Burns, Prof Geoffrey P. Margison**, **Vicent Pelechano**, and **Jingwen Wang** for their contributions. I really appreciate help from **Yerma Pareja** who has been my guide in the collaborator's lab for the RNA-seq project.

In addition, I sincerely thank to all my current and previous colleagues (Ilona, Åse, Ian, Aram, Beatrice) at IMM that supported me during these (almost 5) years of my PhD and made this journey full of support, interesting activities and good times. Thanks, **Emma**, for being dear friend and roommate. We were a perfect team of complaint! Thanks for all the nice memories from our activities. Dear **Katha**, you are our party diva! Without you, it would have been so boring at IMM. Thanks for organizing sauna parties, skiing trips, dinner evenings, Eurovision parties, and much more. Best of the luck with your thesis writing! **Imran**, thanks for our fruitful conversations at the dinner table about future food capsules and other interesting topics. Thanks for your support and time listening about my problems and giving me best advices. **Jessika**, our sweet Jessika, I am so proud of your improvement here and, as we discussed many times, you will do it! (In addition, thanks for spoiling us with home baked cakes). **Sarah** thanks for spreading optimism and good energy as well as always trying to bring up a positive side even at very down moments. **Hanna** thank you for taking the responsibility to be a chairwoman at my defense. **Virginia** thanks for a good vibe at the work place (P.S. your pastries from Portugal were deeply appreciated and already missed so much). **Kristin** thanks for being my roomie and wisely showing me that problems are multi-

sided and other people might see them differently than do I. **Divya,** a special thanks for you. Your urgent help with FACS saved me! **Rong Rong**, wish you a good luck with your PhD studies! **Huiyuan**, wish you all the best with your work. A special thanks to a previous colleague, post doc **João**, for our collaboration that resulted in a nice publication. **Shuo**, wish you a good luck with the new position! **Felipe**, welcome back to Sweden and to our group, wish you many nice results. Best of luck **Anna**, **Anda**, **Jeremy**, **Nanna, Noora, Mizan**, **Ayman**, and **Jessika De Loma**!

I want to thank my family and friends outside the world of research, who (most of them) does not have a clue what I am doing and what does it mean to be "doktorand" in the area of molecular biology. However, it is always nice to have someone around to do funny things that your miserable life of PhD does not have to spin continuously around the research and not working experiments. My dear **Ignas**, I really appreciate your constant support during my PhD studies, all surprise visits with coffee or flowers that would bring sunshine to my day. Thanks for not complaining about my weekends at work and this last month with me out in space, but encouraging me to move on.

The work in this thesis was performed at the Unit of Biochemical Toxicology, Institute of Environmental Medicine, Karolinska Institutet. I would like acknowledge all funding that made my studies possible. I really hope that my findings will lead somewhere and will benefit society in general.

8 REFERENCES

Abbott, P.J., and R. Saffhill. 1979. DNA synthesis with methylated poly(dC-dG) templates. Evidence for a competitive nature to miscoding by O6-methylguanine. *BBA Section Nucleic Acids And Protein Synthesis*. 562:51-61.

Acevedo, A., and R. Andino. 2014. Library preparation for highly accurate population sequencing of RNA viruses. *Nature protocols*. 9:1760-1769.

Acevedo, A., L. Brodsky, and R. Andino. 2014. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature*. 505:686-690.

Anindya, R., P.-O. Mari, U. Kristensen, H. Kool, G. Giglia-Mari, L.H. Mullenders, M. Fousteri, W. Vermeulen, J.-M. Egly, and J.Q. Svejstrup. 2010. A Ubiquitin-Binding Domain in Cockayne Syndrome B Required for Transcription-Coupled Nucleotide Excision Repair. *Molecular cell*. 38:637-648.

Bakke, A.C. 2001. The Principles of Flow Cytometry. *Laboratory Medicine*. 32:207-211.

Basu, S., G. Je, and Y.S. Kim. 2015. Transcriptional mutagenesis by 8-oxodG in alpha-synuclein aggregation and the pathogenesis of Parkinson's disease. *Experimental & molecular medicine*. 47:e179.

Beerens, N., J.H. Hoeijmakers, R. Kanaar, W. Vermeulen, and C. Wyman. 2005. The CSB protein actively wraps DNA. *The Journal of biological chemistry*. 280:4722-4729.

Bielas, J.H., and L.A. Loeb. 2005. Mutator phenotype in cancer: timing and perspectives. *Environmental and molecular mutagenesis*. 45:206-213.

Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual review of biochemistry*. 72:291-336.

Bregeon, D., Z.A. Doddridge, H.J. You, B. Weiss, and P.W. Doetsch. 2003. Transcriptional mutagenesis induced by uracil and 8-oxoguanine in Escherichia coli. *Molecular cell*. 12:959-970.

Bregeon, D., and P.W. Doetsch. 2004. Reliable method for generating double-stranded DNA vectors containing site-specific base modifications. *BioTechniques*. 37:760-762, 764, 766.

Brégeon, D., and P.W. Doetsch. 2011. Transcriptional mutagenesis: causes and involvement in tumor development. *Nature reviews. Cancer*. 11:218-227.

Bregeon, D., P.A. Peignon, and A. Sarasin. 2009. Transcriptional mutagenesis induced by 8-oxoguanine in mammalian cells. *PLoS genetics*. 5:e1000577.

Brody, Y., and Y. Shav-Tal. 2011. Transcription and splicing: When the twain meet. *Transcription*. 2:216-220.

Brulliard, M., D. Lorphelin, O. Collignon, W. Lorphelin, B. Thouvenot, E. Gothie, S. Jacquenet, V. Ogier, O. Roitel, J.M. Monnez, P. Vallois, F.T. Yen, O. Poch, M. Guenneugues, G. Karcher, P. Oudet, and B.E. Bihain. 2007. Nonrandom variations in human cancer ESTs indicate that mRNA heterogeneity increases during carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 104:7522-7527.

Bruzel, A., and V.G. Cheung. 2006. DNA reassociation using oscillating phenol emulsions. *Genomics*. 87:286-289.

Bullock, A.N., and A.R. Fersht. 2001. Rescuing the function of mutant p53. *Nat Rev Cancer*. 1:68-76.

Burley, S.K., and R.G. Roeder. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annual review of biochemistry*. 65:769-799.

Burns, J.A., K. Dreij, L. Cartularo, and D.A. Scicchitano. 2010. O6-methylguanine induces altered proteins at the level of transcription in human cells. *Nucleic Acids Res*. 38:8178-8187.

Carey, L.B. 2015. RNA polymerase errors cause splicing defects and can be regulated by differential expression of RNA polymerase subunits. *eLife*. 4:e09945.

Carstens, R.P., J.V. Eaton, H.R. Krigman, P.J. Walther, and M.A. Garcia-Blanco. 1997. Alternative splicing of fibroblast growth factor receptor 2 (FGF-R2) in human prostate cancer. *Oncogene*. 15:3059-3065.

Cascino, I., G. Fiucci, G. Papoff, and G. Ruberti. 1995. Three functional soluble forms of the human apoptosis-inducing Fas molecule are produced by alternative splicing. *Journal of immunology (Baltimore, Md. : 1950)*. 154:2706-2713.

Charlet-Berguerand, N., S. Feuerhahn, S.E. Kong, H. Ziserman, J.W. Conaway, R. Conaway, and J.M. Egly. 2006. RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. *Embo j*. 25:5481-5491.

Chen, Y.H., and D.F. Bogenhagen. 1993. Effects of DNA lesions on transcription elongation by T7 RNA polymerase. *The Journal of biological chemistry*. 268:5849-5855.

Cheng, K.C., D.S. Cahill, H. Kasai, S. Nishimura, and L.A. Loeb. 1992. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. *The Journal of biological chemistry*. 267:166-172.

Chiara, M.D., and R. Reed. 1995. A two-step mechanism for 5' and 3' splice-site pairing. *Nature*. 375:510-513.

Christmann, M., B. Verbeek, W.P. Roos, and B. Kaina. 2011. O(6)-Methylguanine-DNA methyltransferase (MGMT) in normal tissues and tumors: enzyme activity, promoter methylation and immunohistochemistry. *Biochimica et biophysica acta*. 1816:179-190.

Citterio, E., V. Van Den Boom, G. Schnitzler, R. Kanaar, E. Bonte, R.E. Kingston, J.H. Hoeijmakers, and W. Vermeulen. 2000. ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. *Molecular and cellular biology*. 20:7643-7653.

Clauson, C.L., K.J. Oestreich, J.W. Austin, and P.W. Doetsch. 2010. Abasic sites and strand breaks in DNA cause transcriptional mutagenesis in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*. 107:3657-3662.

Connelly, S., and J.L. Manley. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & development*. 2:440-452.

Cooper, T.A., L. Wan, and G. Dreyfuss. 2009. RNA and disease. *Cell*. 136:777-793.

CST. 2012. Caspase-3 Activity Assay Kit. Cell Signaling Technology, accessed Oct 2018 from: https://media.cellsignal.com/pdf/5723.pdf.

Damsma, G.E., and P. Cramer. 2009. Molecular basis of transcriptional mutagenesis at 8-oxoguanine. *The Journal of biological chemistry*. 284:31658-31663.

Davies, B.S., L.G. Fong, S.H. Yang, C. Coffinier, and S.G. Young. 2009. The posttranslational processing of prelamin A and disease. *Annual review of genomics and human genetics*. 10:153-174.

Davies, R.J., R. Miller, and N. Coleman. 2005. Colorectal cancer screening: prospects for molecular stool analysis. *Nat Rev Cancer*. 5:199-209.

De Bont, R., and N. van Larebeke. 2004. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*. 19:169-185.

Derheimer, F.A., H.M. O'Hagan, H.M. Krueger, S. Hanasoge, M.T. Paulsen, and M. Ljungman. 2007. RPA and ATR link transcriptional stress to p53. *Proceedings of the National Academy of Sciences of the United States of America*. 104:12778-12783.

Dimitri, A., J.A. Burns, S. Broyde, and D.A. Scicchitano. 2008. Transcription elongation past O(6)-methylguanine by human RNA polymerase II and bacteriophage T7 RNA polymerase. *Nucleic Acids Research*. 36:6459-6471.

Doetsch, P.W. 2002. Translesion synthesis by RNA polymerases: occurrence and biological implications for transcriptional mutagenesis. *Mutation research*. 510:131-140.

Dreij, K., J.A. Burns, A. Dimitri, L. Nirenstein, T. Noujnykh, and D.A. Scicchitano. 2010. DNA Damage and Transcription Elongation: Consequences and RNA Integrity. *In* The Chemical Biology of DNA Damage (eds N. E. Geacintov and S. Broyde).

Dutta, D., K. Shatalin, V. Epshtein, M.E. Gottesman, and E. Nudler. 2011. Linking RNA polymerase backtracking to genome instability in E. coli. *Cell*. 146:533-543.

Ebright, R.H. 2000. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *Journal of molecular biology*. 304:687-698.

Ehara, H., T. Yokoyama, H. Shigematsu, S. Yokoyama, M. Shirouzu, and S.I. Sekine. 2017. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science (New York, N.Y.)*. 357:921-924.

el-Deiry, W.S., J.W. Harper, P.M. O'Connor, V.E. Velculescu, C.E. Canman, J. Jackman, J.A. Pietenpol, M. Burrell, D.E. Hill, Y. Wang, and et al. 1994. WAF1/CIP1 is induced in p53-mediated G1 arrest and apoptosis. *Cancer research*. 54:1169-1174.

Eriksson, M., W.T. Brown, L.B. Gordon, M.W. Glynn, J. Singer, L. Scott, M.R. Erdos, C.M. Robbins, T.Y. Moses, P. Berglund, A. Dutra, E. Pak, S. Durkin, A.B. Csoka, M. Boehnke, T.W. Glover, and F.S. Collins. 2003. Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature*. 423:293-298.

Ezerskyte, M., J.A. Paredes, S. Malvezzi, J.A. Burns, G.P. Margison, M. Olsson, D.A. Scicchitano, and K. Dreij. 2018. O(6)-methylguanine-induced transcriptional mutagenesis reduces p53 tumor-suppressor function. *Proceedings of the National Academy of Sciences of the United States of America*. 115:4731-4736.

Fei, J., and J. Chen. 2012. KIAA1530 protein is recruited by Cockayne syndrome complementation group protein A (CSA) to participate in transcription-coupled repair (TCR). *The Journal of biological chemistry*. 287:35118-35126.

Fischer, M. 2017. Census and evaluation of p53 target genes. *Oncogene*. 36:3943-3956.

Fleischer, T., H. Edvardsen, H.K. Solvang, C. Daviaud, B. Naume, A.L. Borresen-Dale, V.N. Kristensen, and J. Tost. 2014. Integrated analysis of high-resolution DNA methylation profiles, gene expression, germline genotypes and clinical end points in breast cancer patients. *International journal of cancer*. 134:2615-2625.

Fousteri, M., W. Vermeulen, A.A. van Zeeland, and L.H. Mullenders. 2006. Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. *Molecular cell*. 23:471-482.

Fox-Walsh, K.L., and K.J. Hertel. 2009. Splice-site pairing is an intrinsically high fidelity process. *Proceedings of the National Academy of Sciences of the United States of America*. 106:1766-1771.

Gamba, P., and N. Zenkin. 2018. Transcription fidelity and its roles in the cell. *Curr Opin Microbiol*. 42:13-18.

Glassner, B.J., G. Weeda, J.M. Allan, J.L. Broekhof, N.H. Carls, I. Donker, B.P. Engelward, R.J. Hampson, R. Hersmus, M.J. Hickman, R.B. Roth, H.B. Warren, M.M. Wu, J.H. Hoeijmakers, and L.D. Samson. 1999. DNA repair methyltransferase (Mgmt) knockout mice are sensitive to the lethal effects of chemotherapeutic alkylating agents. *Mutagenesis*. 14:339-347.

Gordon, A.J.E., D. Satory, J.A. Halliday, and C. Herman. 2015. Lost in transcription: transient errors in information transfer. *Current Opinion in Microbiology*. 24:80-87.

Gout, J.-F., W.K. Thomas, Z. Smith, K. Okamoto, and M. Lynch. 2013. Large-scale detection of in vivo transcription errors. *Proceedings of the National Academy of Sciences*. 110:18584-18589.

Gout, J.F., W. Li, C. Fritsch, A. Li, S. Haroon, L. Singh, D. Hua, H. Fazelinia, Z. Smith, S. Seeholzer, K. Thomas, M. Lynch, and M. Vermulst. 2017. The landscape of transcription errors in eukaryotic cells. *Science advances*. 3:e1701484.

Graham, F.L., J. Smiley, W.C. Russell, and R. Nairn. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *The Journal of general virology*. 36:59-74.

Groisman, R., I. Kuraoka, O. Chevallier, N. Gaye, T. Magnaldo, K. Tanaka, A.F. Kisselev, A. Harel-Bellan, and Y. Nakatani. 2006. CSA-dependent degradation of CSB by the ubiquitin-proteasome pathway establishes a link between complementation factors of the Cockayne syndrome. *Genes & development*. 20:1429-1434.

Groisman, R., J. Polanowska, I. Kuraoka, J. Sawada, M. Saijo, R. Drapkin, A.F. Kisselev, K. Tanaka, and Y. Nakatani. 2003. The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell*. 113:357-367.

Halford, S., A. Rowan, E. Sawyer, I. Talbot, and I. Tomlinson. 2005. O(6)-methylguanine methyltransferase in colorectal cancers: detection of mutations, loss of expression, and weak association with G:C>A:T transitions. *Gut*. 54:797-802.

Hanahan, D., and R.A. Weinberg. 2011. Hallmarks of cancer: the next generation. *Cell*. 144:646-674.

Hanawalt, P.C. 2002. Subpathways of nucleotide excision repair and their regulation. *Oncogene*. 21:8949-8956.

Hanawalt, P.C., and G. Spivak. 2008. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol*. 9:958-970.

Hartl, U.F. 2016. Cellular Homeostasis and Aging. *Annual review of biochemistry*. 85:1-4.

Hastings, M.L., and A.R. Krainer. 2001. Pre-mRNA splicing in the new millennium. *Current opinion in cell biology*. 13:302-309.

Hayflick, L. 1965. THE LIMITED IN VITRO LIFETIME OF HUMAN DIPLOID CELL STRAINS. *Experimental cell research*. 37:614-636.

Hemminki, K. 1983. Nucleic acid adducts of chemical carcinogens and mutagens. *Archives of Toxicology*. 52:249-285.

Henning, K.A., L. Li, N. Iyer, L.D. McDaniel, M.S. Reagan, R. Legerski, R.A. Schultz, M. Stefanini, A.R. Lehmann, L.V. Mayne, and E.C. Friedberg. 1995. The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH. *Cell*. 82:555-564.

Hobson, G.M., Z. Huang, K. Sperle, E. Sistermans, P.K. Rogan, J.Y. Garbern, E. Kolodny, S. Naidu, and F. Cambi. 2006. Splice-site contribution in alternative splicing of PLP1 and DM20: molecular studies in oligodendrocytes. *Human mutation*. 27:69-77.

Horn, H.F., and K.H. Vousden. 2007. Coping with stress: multiple ways to activate p53. *Oncogene*. 26:1306.

House, A.E., and K.W. Lynch. 2008. Regulation of alternative splicing: more than just the ABCs. *The Journal of biological chemistry*. 283:1217-1221.

Huang, J.C., D.S. Hsu, A. Kazantsev, and A. Sancar. 1994. Substrate spectrum of human excinuclease: repair of abasic sites, methylated bases, mismatches, and bulky adducts. *Proceedings of the National Academy of Sciences of the United States of America*. 91:12213-12217.

Illumina. 2017. An introduction to Next-Generation Sequencing Technology, accessed Oct 2018 from: https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf .

Imashimizu, M., T. Oshima, L. Lubkowska, and M. Kashlev. 2013. Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Res*. 41:9090-9104.

Itoh, T., T. Ono, and M. Yamaizumi. 1994. A new UV-sensitive syndrome not belonging to any complementation groups of xeroderma pigmentosum or Cockayne syndrome: siblings showing biochemical characteristics of Cockayne syndrome without typical clinical manifestations. *Mutation research*. 314:233-248.

Jiang, Z., J. Hu, X. Li, Y. Jiang, W. Zhou, and D. Lu. 2006. Expression analyses of 27 DNA repair genes in astrocytoma by TaqMan low-density array. *Neuroscience letters*. 409:112-117.

Jonkers, I., and J.T. Lis. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*. 16:167.

Juven-Gershon, T., and J.T. Kadonaga. 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental biology*. 339:225-229.

Kaina, B., M. Christmann, S. Naumann, and W.P. Roos. 2007. MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA repair*. 6:1079-1099.

Kim, T.K., R.H. Ebright, and D. Reinberg. 2000. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science (New York, N.Y.)*. 288:1418-1422.

Klungland, A., I. Rosewell, S. Hollenbach, E. Larsen, G. Daly, B. Epe, E. Seeberg, T. Lindahl, and D.E. Barnes. 1999. Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. *Proceedings of the National Academy of Sciences of the United States of America*. 96:13300-13305.

Knippa, K., and D.O. Peterson. 2013. Fidelity of RNA polymerase II transcription: Role of Rpb9 [corrected] in error detection and proofreading. *Biochemistry*. 52:7807-7817.

Kramer, W., and H.J. Fritz. 1987. Oligonucleotide-directed construction of mutations via gapped duplex DNA. *Methods in enzymology*. 154:350-367.

Krishan, A. 1975. Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining. *The Journal of cell biology*. 66:188-193.

Krokan, H.E., R. Standal, and G. Slupphaug. 1997. DNA glycosylases in the base excision repair of DNA. *The Biochemical journal*. 325 (Pt 1):1-16.

Kukurba, K.R., and S.B. Montgomery. 2015. RNA Sequencing and Analysis. *Cold Spring Harbor protocols*. 2015:951-969.

Kulaeva, O.I., F.K. Hsieh, H.W. Chang, D.S. Luse, and V.M. Studitsky. 2013. Mechanism of Transcription through a Nucleosome by RNA polymerase II. *Biochimica et biophysica acta*. 1829:76-83.

Kuraoka, I., M. Endou, Y. Yamaguchi, T. Wada, H. Handa, and K. Tanaka. 2003. Effects of endogenous DNA base lesions on transcription elongation by mammalian RNA polymerase II. Implications for transcription-coupled DNA repair and transcriptional mutagenesis. *The Journal of biological chemistry*. 278:7294-7299.

Lagerwerf, S., M.G. Vrouwe, R.M. Overmeer, M.I. Fousteri, and L.H. Mullenders. 2011. DNA damage response and transcription. *DNA repair*. 10:743-750.

Lancaster, M.V., and R.D. Fields. 1996. Antibiotic And Cytotoxic Drug Susceptibility Assays Using Resazurin And Poising Agents. Patent Number US 5,501,959.

Lewin, B., J. Krebs, S.T. Kilpatrick, and E.S. Goldstein. 2011. Lewin's GENES X. Jones & Bartlett Learning.

Lindahl, T. 1993. Instability and decay of the primary structure of DNA. *Nature*. 362:709-715.

Lindsey-Boltz, L.A., and A. Sancar. 2007. RNA polymerase: The most specific damage recognition protein in cellular responses to DNA damage? *Proceedings of the National Academy of Sciences of the United States of America*. 104:13213-13214.

Liu, J., W. Zhou, and P.W. Doetsch. 1995. RNA polymerase bypass at sites of dihydrouracil: implications for transcriptional mutagenesis. *Molecular and cellular biology*. 15:6729-6735.

Ljungman, M., and D.P. Lane. 2004. Transcription — guarding the genome by sensing DNA damage. *Nature Reviews Cancer*. 4:727.

Loeb, K.R., and L.A. Loeb. 2000. Significance of multiple mutations in cancer. *Carcinogenesis*. 21:379-385.

Logan, J., E. Falck-Pedersen, J.E. Darnell, and T. Shenk. 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proceedings of the National Academy of Sciences of the United States of America*. 84:8306-8310.

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. 15:550.

Luhnsdorf, B., N. Kitsera, D. Warken, T. Lingg, B. Epe, and A. Khobta. 2012. Generation of reporter plasmids containing defined base modifications in the DNA strand of choice. *Analytical biochemistry*. 425:47-53.

Luse, D.S. 2013. Promoter clearance by RNA polymerase II. *Biochimica et biophysica acta*. 1829:63-68.

Lynch, M. 2010. Evolution of the mutation rate. *Trends in Genetics*. 26:345-352.

Margison, G.P., A. Butt, S.J. Pearson, S. Wharton, A.J. Watson, A. Marriott, C.M. Caetano, J.J. Hollins, N. Rukazenkova, G. Begum, and M.F. Santibanez-Koref. 2007. Alkyltransferase-like proteins. *DNA repair*. 6:1222-1228.

Margison, G.P., and M.F. Santibanez-Koref. 2002. O6-alkylguanine-DNA alkyltransferase: role in carcinogenesis and chemotherapy. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 24:255-266.

Marietta, C., and P.J. Brooks. 2007. Transcriptional bypass of bulky DNA lesions causes new mutant RNA transcripts in human cells. *EMBO reports*. 8:388-393.

Markowitz, S.D., and M.M. Bertagnolli. 2009. Molecular Basis of Colorectal Cancer. *New England Journal of Medicine*. 361:2449-2460.

Matlin, A.J., F. Clark, and C.W.J. Smith. 2005. Understanding alternative splicing: towards a cellular code. *Nature Reviews Molecular Cell Biology*. 6:386.

Mellon, I., G. Spivak, and P.C. Hanawalt. 1987. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell*. 51:241-249.

Mishina, Y., E.M. Duguid, and C. He. 2006. Direct reversal of DNA alkylation damage. *Chemical reviews*. 106:215-232.

Moore, M.J. 2000. Intron recognition comes of AGe. *Nature structural biology*. 7:14-16.

Morreall, J., A. Kim, Y. Liu, N. Degtyareva, B. Weiss, and P.W. Doetsch. 2015. Evidence for Retromutagenesis as a Mechanism for Adaptive Mutation in Escherichia coli. *PLoS genetics*. 11:e1005477.

Morreall, J.F., L. Petrova, and P.W. Doetsch. 2013. Transcriptional mutagenesis and its potential roles in the etiology of cancer and bacterial antibiotic resistance. *Journal of cellular physiology*. 228:2257-2261.

Mullenders, L. 2015. DNA damage mediated transcription arrest: Step back to go forward. *DNA repair*. 36:28-35.

Nagel, Z.D., C.M. Margulies, I.A. Chaim, S.K. McRee, P. Mazzucato, A. Ahmad, R.P. Abo, V.L. Butty, A.L. Forget, and L.D. Samson. 2014. Multiplexed DNA repair assays for

multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 111:E1823-1832.

Nakazawa, Y., K. Sasaki, N. Mitsutake, M. Matsuse, M. Shimada, T. Nardo, Y. Takahashi, K. Ohyama, K. Ito, H. Mishima, M. Nomura, A. Kinoshita, S. Ono, K. Takenaka, R. Masuyama, T. Kudo, H. Slor, A. Utani, S. Tateishi, S. Yamashita, M. Stefanini, A.R. Lehmann, K.-i. Yoshiura, and T. Ogi. 2012. Mutations in UVSSA cause UV-sensitive syndrome and impair RNA polymerase IIo processing in transcription-coupled nucleotide-excision repair. *Nature Genetics*. 44:586.

Nance, M.A., S. Boyadjiev, V.M. Pratt, S. Taylor, M.E. Hodes, and S.R. Dlouhy. 1996. Adult-onset neurodegenerative disorder due to proteolipid protein gene mutation in the mother of a man with Pelizaeus-Merzbacher disease. *Neurology*. 47:1333-1335.

Nesser, N.K., D.O. Peterson, and D.K. Hawley. 2006. RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. *Proceedings of the National Academy of Sciences of the United States of America*. 103:3268-3273.

Neumann, E., M. Schaefer-Ridder, Y. Wang, and P.H. Hofschneider. 1982. Gene transfer into mouse lyoma cells by electroporation in high electric fields. *Embo j*. 1:841-845.

Nicholls, D.G., and M.W. Ward. 2000. Mitochondrial membrane potential and neuronal glutamate excitotoxicity: mortality and millivolts. *Trends in neurosciences*. 23:166-174.

Nilsen, T.W., and B.R. Graveley. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature*. 463:457.

Nogales, E., R.K. Louder, and Y. He. 2017. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annual review of biophysics*. 46:59-83.

Nouspikel, T. 2009. DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. *Cellular and molecular life sciences : CMLS*. 66:994-1009.

Nouspikel, T., and P.C. Hanawalt. 2002. DNA repair in terminally differentiated cells. *DNA repair*. 1:59-75.

O'Farrell, T.J., P. Ghosh, N. Dobashi, C.Y. Sasaki, and D.L. Longo. 2004. Comparison of the effect of mutant and wild-type p53 on global gene expression. *Cancer research*. 64:8199-8207.

Orphanides, G., T. Lagrange, and D. Reinberg. 1996. The general transcription factors of RNA polymerase II. *Genes & development*. 10:2657-2683.

Pack, D.W., A.S. Hoffman, S. Pun, and P.S. Stayton. 2005. Design and development of polymers for gene delivery. *Nature reviews. Drug discovery*. 4:581-593.

Pan, Q., O. Shai, L.J. Lee, B.J. Frey, and B.J. Blencowe. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*. 40:1413.

Pandit, S., D. Wang, and X.D. Fu. 2008. Functional integration of transcriptional and RNA processing machineries. *Current opinion in cell biology*. 20:260-265.

Paredes, J.A., M. Ezerskyte, M. Bottai, and K. Dreij. 2017. Transcriptional mutagenesis reduces splicing fidelity in mammalian cells. *Nucleic Acids Res*. 45:6520-6529.

Pastoriza-Gallego, M., J. Armier, and A. Sarasin. 2007. Transcription through 8-oxoguanine in DNA repair-proficient and Csb(-)/Ogg1(-) DNA repair-deficient mouse embryonic fibroblasts is dependent upon promoter strength and sequence context. *Mutagenesis*. 22:343-351.

Perry, S.W., J.P. Norman, J. Barbieri, E.B. Brown, and H.A. Gelbard. 2011. Mitochondrial membrane potential probes and the proton gradient: a practical usage guide. *BioTechniques*. 50:98-115.

Petitjean, A., M.I. Achatz, A.L. Borresen-Dale, P. Hainaut, and M. Olivier. 2007. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene*. 26:2157-2165.

Proudfoot, N.J., A. Furger, and M.J. Dye. 2002. Integrating mRNA processing with transcription. *Cell*. 108:501-512.

Rampersad, S.N. 2012. Multiple applications of Alamar Blue as an indicator of metabolic function and cellular health in cell viability bioassays. *Sensors (Basel, Switzerland)*. 12:12347-12360.

Rapin, I., Y. Lindenbaum, D.W. Dickson, K.H. Kraemer, and J.H. Robbins. 2000. Cockayne syndrome and xeroderma pigmentosum. *Neurology*. 55:1442-1449.

Reed, R., and T. Maniatis. 1988. The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes & development*. 2:1268-1276.

Reid-Bayliss, K.S., and L.A. Loeb. 2017. Accurate RNA consensus sequencing for high-fidelity detection of transcriptional mutagenesis-induced epimutations. *Proceedings of the National Academy of Sciences of the United States of America*. 114:9415-9420.

Resnick, M.A., and A. Inga. 2003. Functional mutants of the sequence-specific transcription factor p53 and implications for master genes of diversity. *Proceedings of the National Academy of Sciences of the United States of America*. 100:9934-9939.

Richard, P., and J.L. Manley. 2009. Transcription termination by nuclear RNA polymerases. *Genes & development*. 23:1247-1269.

Rino, J., and M. Carmo-Fonseca. 2009. The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends in cell biology*. 19:375-384.

Roeder, R.G. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci*. 21:327-335.

Rydberg, B., and T. Lindahl. 1982. Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *The EMBO Journal*. 1:211-216.

Sakumi, K., Y. Tominaga, M. Furuichi, P. Xu, T. Tsuzuki, M. Sekiguchi, and Y. Nakabeppu. 2003. Ogg1 knockout-associated lung tumorigenesis and its suppression by Mth1 gene disruption. *Cancer research*. 63:902-905.

Sassanfar, M., M.K. Dosanjh, J.M. Essigmann, and L. Samson. 1991. Relative efficiencies of the bacterial, yeast, and human DNA methyltransferases for the repair of O6-methylguanine and O4-methylthymine. Suggestive evidence for O4-methylthymine repair by eukaryotic methyltransferases. *The Journal of biological chemistry*. 266:2767-2771.

Saxowsky, T.T., and P.W. Doetsch. 2006. RNA polymerase encounters with DNA damage: transcription-coupled repair or transcriptional mutagenesis? *Chemical reviews*. 106:474-488.

Saxowsky, T.T., K.L. Meadows, A. Klungland, and P.W. Doetsch. 2008. 8-Oxoguanine-mediated transcriptional mutagenesis causes Ras activation in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*. 105:18877-18882.

Schellenberg, M.J., D.B. Ritchie, and A.M. MacMillan. 2008. Pre-mRNA splicing: a complex picture in higher definition. *Trends Biochem Sci*. 33:243-246.

Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. 2011. Global quantification of mammalian gene expression control. *Nature*. 473:337-342.

Schwertman, P., A. Lagarou, D.H. Dekkers, A. Raams, A.C. van der Hoek, C. Laffeber, J.H. Hoeijmakers, J.A. Demmers, M. Fousteri, W. Vermeulen, and J.A. Marteijn. 2012. UV-sensitive syndrome protein UVSSA recruits USP7 to regulate transcription-coupled repair. *Nat Genet*. 44:598-602.

Scian, M.J., K.E. Stagliano, M.A. Ellis, S. Hassan, M. Bowman, M.F. Miles, S.P. Deb, and S. Deb. 2004. Modulation of gene expression by tumor-derived p53 mutants. *Cancer research*. 64:7447-7454.

Selby, C.P., R. Drapkin, D. Reinberg, and A. Sancar. 1997. RNA polymerase II stalled at a thymine dimer: footprint and effect on excision repair. *Nucleic Acids Res*. 25:787-793.

Sharp, P.A. 1987. Splicing of messenger RNA precursors. *Science (New York, N.Y.)*. 235:766-771.

Shaw, R.J., N.D. Bonawitz, and D. Reines. 2002. Use of an in vivo reporter assay to test for transcriptional and translational fidelity in yeast. *The Journal of biological chemistry*. 277:24420-24426.

Shibutani, S., M. Takeshita, and A.P. Grollman. 1991. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature*. 349:431-434.

Sims, R.J., 3rd, R. Belotserkovskaya, and D. Reinberg. 2004. Elongation by RNA polymerase II: the short and long of it. *Genes & development*. 18:2437-2468.

Singh, J., L. Su, and E.T. Snow. 1996. Replication across O6-methylguanine by human DNA polymerase beta in vitro. Insights into the futile cytotoxic repair and mutagenesis of O6-methylguanine. *The Journal of biological chemistry*. 271:28391-28398.

Spivak, G. 2016. Transcription-coupled repair: an update. *Arch Toxicol*. 90:2583-2594.

Stein, L.D. 2004. Human genome: end of the beginning. *Nature*. 431:915-916.

Strano, S., S. Dell'Orso, S. Di Agostino, G. Fontemaggi, A. Sacchi, and G. Blandino. 2007. Mutant p53: an oncogenic transcription factor. *Oncogene*. 26:2212-2219.

Suñé-Pou, M., S. Prieto-Sánchez, S. Boyero-Corral, C. Moreno-Castro, Y. El Yousfi, J.M. Suñé-Negre, C. Hernández-Munain, and C. Suñé. 2017. Targeting Splicing in the Treatment of Human Disease. *Genes*. 8.

Sveen, A., S. Kilpinen, A. Ruusulehto, R.A. Lothe, and R.I. Skotheim. 2016. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene*. 35:2413-2427.

Svejstrup, J.Q. 2002. Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell Biol*. 3:21-29.

Svejstrup, J.Q. 2004. The RNA polymerase II transcription cycle: cycling through chromatin. *Biochimica et biophysica acta*. 1677:64-73.

Svejstrup, J.Q. 2007. Contending with transcriptional arrest during RNAPII transcript elongation. *Trends Biochem Sci*. 32:165-171.

Svejstrup, J.Q. 2010. The interface between transcription and mechanisms maintaining genome integrity. *Trends in Biochemical Sciences*. 35:333-338.

Sydow, J.F., and P. Cramer. 2009. RNA polymerase fidelity and transcriptional proofreading. *Current opinion in structural biology*. 19:732-739.

Tano, K., S. Shiota, J. Collier, R.S. Foote, and S. Mitra. 1990. Isolation and structural characterization of a cDNA clone encoding the human DNA repair protein for O6-alkylguanine. *Proceedings of the National Academy of Sciences of the United States of America*. 87:686-690.

Tchou, J., H. Kasai, S. Shibutani, M.H. Chung, J. Laval, A.P. Grollman, and S. Nishimura. 1991. 8-oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity. *Proceedings of the National Academy of Sciences of the United States of America*. 88:4690-4694.

Tisdale, M.J. 1987. Antitumor imidazotetrazines--XV. Role of guanine O6 alkylation in the mechanism of cytotoxicity of imidazotetrazinones. *Biochem Pharmacol*. 36:457-462.

Tornaletti, S., B.A. Donahue, D. Reines, and P.C. Hanawalt. 1997. Nucleotide sequence context effect of a cyclobutane pyrimidine dimer upon RNA polymerase II transcription. *The Journal of biological chemistry*. 272:31719-31724.

Tornaletti, S., and P.C. Hanawalt. 1999. Effect of DNA lesions on transcription elongation. *Biochimie*. 81:139-146.

Traverse, C.C., and H. Ochman. 2016. Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. *Proceedings of the National Academy of Sciences of the United States of America*. 113:3311-3316.

Tubbs, J.L., V. Latypov, S. Kanugula, A. Butt, M. Melikishvili, R. Kraehenbuehl, O. Fleck, A. Marriott, A.J. Watson, B. Verbeek, G. McGown, M. Thorncroft, M.F. Santibanez-Koref, C. Millington, A.S. Arvai, M.D. Kroeger, L.A. Peterson, D.M. Williams, M.G. Fried, G.P. Margison, A.E. Pegg, and J.A. Tainer. 2009. Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature*. 459:808-813.

Walmacq, C., L. Wang, J. Chong, K. Scibelli, L. Lubkowska, A. Gnatt, P.J. Brooks, D. Wang, and M. Kashlev. 2015. Mechanism of RNA polymerase II bypass of oxidative cyclopurine DNA lesions. *Proceedings of the National Academy of Sciences of the United States of America*. 112:E410-419.

van den Boom, V., N.G. Jaspers, and W. Vermeulen. 2002. When machines get stuck--obstructed RNA polymerase II: displacement, degradation or suicide. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 24:780-784.

van Leeuwen, F.W., D.P. de Kleijn, H.H. van den Hurk, A. Neubauer, M.A. Sonnemans, J.A. Sluijs, S. Koycu, R.D. Ramdjielal, A. Salehi, G.J. Martens, F.G. Grosveld, J. Peter, H. Burbach, and E.M. Hol. 1998. Frameshift mutants of beta amyloid precursor protein and ubiquitin-B in Alzheimer's and Down patients. *Science (New York, N.Y.)*. 279:242-247.

Wang, E.T., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S.F. Kingsmore, G.P. Schroth, and C.B. Burge. 2008. Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature*. 456:470-476.

Wang, Z., and C.B. Burge. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA (New York, N.Y.)*. 14:802-813.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*. 10:57-63.

Vaquero-Garcia, J., A. Barrera, M.R. Gazzara, J. Gonzalez-Vallinas, N.F. Lahens, J.B. Hogenesch, K.W. Lynch, and Y. Barash. 2016. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife*. 5:e11752.

Ward, A.J., and T.A. Cooper. 2010. The Pathobiology of Splicing. *The Journal of pathology*. 220:152-163.

Watson, A.J., and G.P. Margison. 2000. O6-alkylguanine-DNA alkyltransferase assay. *Methods in molecular biology (Clifton, N.J.)*. 152:49-61.

Wells, P.G., G.P. McCallum, C.S. Chen, J.T. Henderson, C.J. Lee, J. Perstin, T.J. Preston, M.J. Wiley, and A.W. Wong. 2009. Oxidative stress in developmental origins of disease: teratogenesis, neurodevelopmental deficits, and cancer. *Toxicological sciences : an official journal of the Society of Toxicology*. 108:4-18.

Vermulst, M., A.S. Denney, M.J. Lang, C.W. Hung, S. Moore, M.A. Moseley, J.W. Thompson, V. Madden, J. Gauer, K.J. Wolfe, D.W. Summers, J. Schleit, G.L. Sutphin, S. Haroon, A. Holczbauer, J. Caine, J. Jorgenson, D. Cyr, M. Kaeberlein, J.N. Strathern, M.C. Duncan, and D.A. Erie. 2015. Transcription errors induce proteotoxic stress and shorten cellular lifespan. *Nature communications*. 6:8065.

Willis, A., E.J. Jung, T. Wakefield, and X. Chen. 2004. Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. *Oncogene*. 23:2330.

Viswanathan, A., and P.W. Doetsch. 1998. Effects of nonbulky DNA base damages on Escherichia coli RNA polymerase-mediated elongation and promoter clearance. *The Journal of biological chemistry*. 273:21276-21281.

von Hippel, P.H. 1998. An integrated model of the transcription complex in elongation, termination, and editing. *Science (New York, N.Y.)*. 281:660-665.

Woodward, K., and S. Malcolm. 1999. Proteolipid protein gene: Pelizaeus-Merzbacher disease in humans and neurodegeneration in mice. *Trends in genetics : TIG*. 15:125-128.

Xu, L., W. Wang, J. Wu, J.H. Shin, P. Wang, I.C. Unarta, J. Chong, Y. Wang, and D. Wang. 2017. Mechanism of DNA alkylation-induced transcriptional stalling, lesion bypass, and mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 114:E7082-e7091.

Yang, W. 2007. Structure and mechanism for DNA lesion recognition. *Cell research*. 18:184.

Ye, S., S. Dhillon, X. Ke, A.R. Collins, and I.N. Day. 2001. An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res*. 29:E88-88.

You, C., X. Dai, B. Yuan, J. Wang, J. Wang, P.J. Brooks, L.J. Niedernhofer, and Y. Wang. 2012. A quantitative assay for assessing the effects of DNA lesions on transcription. *Nature chemical biology*. 8:817-822.

You, C., J. Wang, X. Dai, and Y. Wang. 2015. Transcriptional inhibition and mutagenesis induced by N-nitroso compound-derived carboxymethylated thymidine adducts in DNA. *Nucleic Acids Res*. 43:1012-1018.

You, H.J., A. Viswanathan, and P.W. Doetsch. 2000. In vivo technique for determining transcriptional mutagenesis. *Methods (San Diego, Calif.)*. 22:120-126.

Yuzenkova, Y., A. Bochkareva, V.R. Tadigotla, M. Roghanian, S. Zorov, K. Severinov, and N. Zenkin. 2010. Stepwise mechanism for transcription fidelity. *BMC Biology*. 8:54.

Zerdoumi, Y., R. Lanos, S. Raad, J.M. Flaman, G. Bougeard, T. Frebourg, and I. Tournier. 2017. Germline TP53 mutations result into a constitutive defect of p53 DNA binding and transcriptional response to DNA damage. *Human molecular genetics*. 26:2591-2602.

Zhou, W., and P.W. Doetsch. 1993. Effects of abasic sites and DNA single-strand breaks on prokaryotic RNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America*. 90:6601-6605.

Zoller, M.J., and M. Smith. 1983. Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. *Methods in enzymology*. 100:468-500.